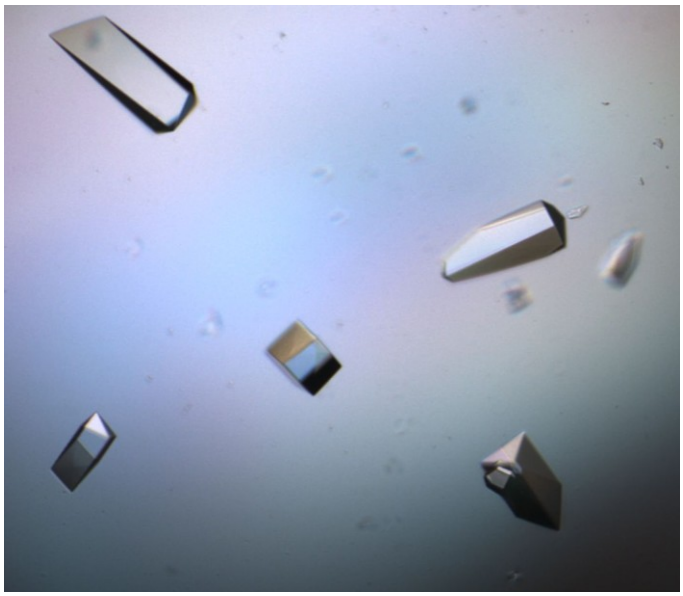
CHALMERS | UNIVERSITY OF GOTHENBURG



Correlation between ligand solubility and formation of protein-ligand complexes in X-ray crystallography

Master of Science Thesis

EMMA JONASSON

Department of Biochemistry, Biomedicine and Biotechnology
UNIVERSITY OF GOTHENBURG
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden, January 2012

Correlation between ligand solubility and formation of protein-ligand complexes in X-ray crystallography

Master of Science Thesis

EMMA JONASSON

Department of Biochemistry, Biomedicine and Biotechnology
UNIVERSITY OF GOTHENBURG
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden, January 2012

Correlation between ligand solubility and formation of protein-ligand complexes in X-ray crystallography

EMMA JONASSON

Master of Science thesis

Department of Biochemistry, Biomedicine and Biotechnology
University of Gothenburg
SE-405 30 Göteborg
Sweden
Telephone: +46 (0)31-786 00 00
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: +46 (0)31-772 1000

Cover: A figure showing crystals obtained in the project, see page 24
Göteborg, Sweden 2012

Correlation between ligand solubility and formation of protein-ligand complexes in X-ray crystallography

EMMA JONASSON
Department of Biochemistry, Biomedicine and Biotechnology
University of Gothenburg
Chalmers University of Technology

Abstract

Structure determination of ligands bound to their target proteins using X-ray crystallography is an important part in the drug discovery process. However, this is not always easy to achieve, especially in cases where the affinity is poor and the solubility of the ligand in the crystallization condition is low. The aim of this project was to investigate a possible correlation between the solubility of a ligand in a soaking condition and the possibility to form protein-ligand complexes by using different cosolvents. Further conclusions were drawn by comparing the results obtained with properties of the ligands.

Two protein systems were used of which protein A was combined with seven different ligands and two isoforms of protein B were combined with one poorly soluble ligand. Two different temperatures, 20°C and 30°C, were used. The solubilities of the ligands in the soaking condition were determined with nuclear magnetic resonance (NMR) spectroscopy and structure determination was performed with X-ray crystallography.

The results showed that the choice of cosolvent affects the solubility in the soaking condition and also that the distribution coefficient, logD, of the ligand could help in deciding with solvent to use. A 10°C difference in temperature did not affect the solubility or the complex formation, based on the results obtained. A tendency could be seen that a higher solubility in the soaking condition increases the possibility to obtain ligand-protein complexes.

Further experiments in the future could help confirm the results from this project and draw new conclusions. Other proteins, ligands and solvents could be used and other parameters could be investigated like pH or a wider difference in temperature.

Keywords: X-ray crystallography, Crystallization, NMR, solubility, cosolvents

Preface

This project is the master thesis for the degree of Master of Science in Biotechnology from Chalmers University of Technology. It has been carried out at AstraZeneca, Mölndal during 20 weeks in the autumn semester 2011.

There are a number of people who have helped making this project possible. First of all I want to thank my supervisors Linda Öster and Jenny Sandmark for all help with planning the project and for always being there to answer my questions and help me along the way.

I also want to thank Per-Olof Eriksson and Gunnar Grönberg for helping me with my NMR studies, Rob Horsefield and Cristian Bodin for providing me with crystals and crystallization information regarding the two systems, Kalle Sigfridsson for sharing knowledge about solvent systems and providing me with nanosuspensions and cyclodextrin solution, Brian Middleton, Emma Evertsson and Marita Olsson for their help with statistics regarding planning of experiments and interpretation of results and Hans-Georg Beisel for providing me with a script to process protein A data.

Finally I want to thank everyone else at AstraZeneca, especially at the Structure & Biophysics department, for all the help I have received and for the good time I have had at the company.

Emma Jonasson
Göteborg, January 2012

Abbreviations

DMSO	Dimethyl sulfoxide
PEG	Polyethylene glycol
DMA	N,N-dimethylacetamide
NMR	Nuclear Magnetic Resonance
TMSP	Trimethylsilyl propionic acid
RF	Radiofrequency
FID	Free-induction decay
K_D	Dissociation constant
logD	Distribution coefficient

Contents

1. Introduction

Structure-aided drug design is an important part of the drug discovery process. Structure determination of a complex of a ligand bound to its target protein with X-ray crystallography is commonly used to develop the ligand or lead compound.

There is a great interest in the pharmaceutical industry to identify drug molecules that target relevant proteins and thereby an interest in improving the methods to find usable leads [1]. High throughput screening (HTS) is a common approach to achieving this, where a large library of compounds can be screened against each target protein. More recently, fragment-based lead discovery (FBLD) has been developed as an alternative strategy, where small, less complex, molecules are found which are then linked or expanded. The advantages with this strategy include screening of a high proportion of the chemical space as well as a higher free energy of binding relative to the number of non-hydrogen atoms. Even though this approach has many advantages there are also some negative aspects including the often low affinity binding of the fragments to the target and possible difficulties in expanding or linking the fragments. Since the binding modes of fragments are difficult to predict, structural information is important for fragment expansion [2].

In crystallography, a low affinity together with a low solubility can lead to problems in obtaining crystal structures of ligand-protein complexes [3]. One possibility to increase the success rate in number of complexes could therefore be to increase the solubility of the ligand in the soaking condition by using different cosolvents. To test this hypothesis, a number of different ligands were in this project combined with different solvents to investigate the effect on solubility. The possibility of obtaining crystal structures of ligand-protein complexes was investigated with X-ray crystallography and the results compared.

1.1. Aim

The aim of the project was to investigate if there is a positive correlation between solubility of a ligand in a soaking condition and the number of crystallized ligand-protein complexes obtained.

1.2. Project description

The questions this project intended to answer were:

- What is the solubility of the studied ligands in the actual soaking conditions?
- Does the choice of cosolvent affect that solubility?
- Can crystal structures be obtained of ligand-protein complexes for the ligands tested?
- Is there a correlation between solubility and success rate in obtaining complexes?
- Do other properties of the ligands correlate to solubility and/or success rate in obtaining complexes?

The scope of the project was to study two proteins. Protein A was combined with seven ligands of which many are small fragments but some larger ligands. The ligands were chosen based on confirmed specific binding to the binding site of the protein as well as being diverse with respect to properties like size and charge. For protein B, two isoforms with similar binding specificity were used. These two proteins were combined with one large ligand with shown poor solubility. Five different solvents were used in the project as well as nanosuspensions for protein B. For project A, two different temperatures were used. The number of conditions and substances tested was planned based on the time limit of the

project. The results achieved during the project also affected the choices of experiments performed, especially for project B. Further, properties were collected for the ligands and compared to the experimental results in order to see trends.

There is a limitation in the information that will be reported in this project. The project is a methodology enhancement project which means that the structures of the ligands as well as the identity of the proteins will be excluded from this report. However, all methods used and conclusions drawn during the project will be fully reported.

2. Background

This section will present the theoretical background to the project including information about solvents and substances used as well as the methods that have been used in this project.

2.1. Solubility and solvents

One of the main tasks of this project was to determine solubility of ligands in different soaking conditions. Here, some information regarding solubility properties used in the project are described as well as some short background to the solvents.

2.1.1. Solubility

Solubility is measured in moles per liter (mol/L) and is the maximal amount of a substance that dissolves in one liter of water. Solubility can be described in different ways including intrinsic solubility, which regards the neutral form of the substance, and apparent solubility, which can be measured for ionizable compounds at a certain pH. Conversion between intrinsic and apparent solubility can be made if the ionization constant, pKa, of the compound is known [4].

The net charge of a molecule at a certain pH can be calculated out of its pKa values using the Hendersson-Hasselbach equation, equation 1, as described by Moore [5].

$$pH = pKa + \log\frac{A^-}{HA} \qquad (1)$$

Charged molecules are generally more water soluble than uncharged molecules, meaning that solubility is affected by the pH of the surrounding solution. The solubility can be enhanced with different strategies. One is to use sonication which is a way of improving solubility and dissolution rate without affecting the stability of the substance [6]. Other methods to increase solubility include heating and the use of cosolvents.

Another property connected to solubility which is often used for drug molecules is the distribution coefficient, logD. LogD is a pH-dependent value which via the pKa value of the substance can be connected to logP, the octanol/water partition coefficient for the molecule in a neutral state [7].

2.1.2. Solvents

The solvents that were used in this project were water, dimethyl sulfoxide (DMSO), cyclodextrin, a mixture of dimethylacetamide (DMA), polyethylene glycol (PEG) 400 and water and also HCl or NaOH depending on ligand structure. Finally, nanosuspensions were made for one ligand. Water was seen as a control to more easily compare the effect of adding a cosolvent.

DMSO, figure 1, is widely used in the pharmaceutical industry and is regarded the most powerful readily available organic solvent. Its solubilizing ability is probably a result of its high dielectric constant but can also be linked to stereochemistry since the molecule is a trigonal pyramid with a lone electron pair on top [4]. DMSO had earlier been used in crystallographic studies with the ligands and was included as a control.

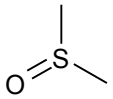


Figure 1. Molecular structure of DMSO.

The remaining solvents have not previously been used for soaking protein crystals in house. The solvents were selected because of their use in other parts of the drug development process, e.g. *in vivo* administration.

Cyclodextrins are substances that can be used to enhance the solubility of insoluble drug molecules, and are used in drug delivery for this purpose. They are cyclic oligosaccharides that consist of six or more glucose units linked with α-(1,4) bonds. β-cyclodextrin with seven glucose units and an inner diameter of 6-6.5 Å, figure 2, is the most commonly used in the pharmaceutical industry. The molecules order so that the primary and secondary hydroxyl groups are situated on either edge of the ring and the hydrophobic groups on the inside [8].

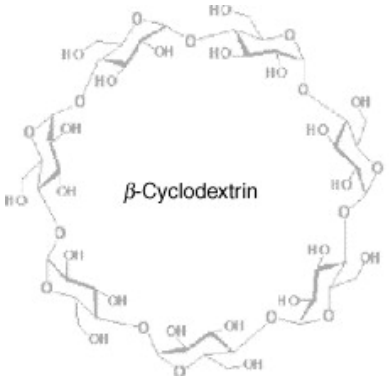


Figure 2. Molecular structure of β-cyclodextrin [8].

By forming inclusion complexes with a molecule, cyclodextrins can alter properties of that molecule, including solubility and stability which has made it useful in the pharmaceutical industry. [9]. Inclusion complexes are formed by binding of non-polar residues of molecules in the hydrophobic cavity of the cyclodextrin molecule. The drive for the binding is mainly the release of water molecules from the cavity. The most common solvent is water but if the molecule dissolves poorly in water the complexation is affected and might be slower or impossible why another solvent might be better. Dissociation of the complex is often achieved by increasing the amount of water and is a relatively fast process [8]. Cyclodextrins have been modified for further improving their properties. One of these modified versions that have shown good results, both regarding performance and safety, is hydroxypropyl- β-cyclodextrin [9]. This is the cyclodextrin variant that has been used in this project.

DMA:PEG400:H2O 1:1:1 has earlier been used successfully to increase solubility [10]. DMA, figure 3, is a water-soluble organic solvent and both DMA and PEG400 can be used to increase the solubility of molecules [11].

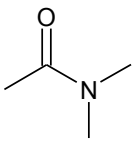


Figure 3. Molecular structure of DMA.

Polyethylene glycols (PEGs), figure 4, are polymers with the chemical formula $HO(CH_2CH_2O)_nH$. The properties and uses of PEGs vary with the weight but general properties like the ability to dissolve in water and organic solvents as well as the absence of toxic effects have made PEGs useful in the biotechnology and medicine industries [12].

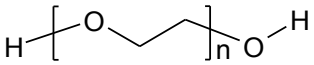


Figure 4. General molecular structure of PEGs.

PEGs are commonly used as precipitants in protein crystallization making the use of them as cosolvents in that kind of experiments even more interesting.

Using HCl or NaOH to more efficiently dissolve the ligands comes from the fact that ionization of basic or acidic components of the molecule often improves the aqueous solubility [6].

One way of formulating poorly soluble drugs is to make nanosuspensions in which the drug powder is milled into nano sized particles. This will increase the surface area resulting in an enhanced dissolution rate and is an interesting method especially for molecules that are poorly soluble both in water and organic solvents [10]. Nanosuspensions can be made from amorphous or crystalline compounds. An amorphous nanosuspension for a compound poorly soluble in water is made by dissolving the substance in an organic solvent that is soluble in water and thereby mixing the solution with an aqueous stabilizer solution. Nanosuspensions from crystalline compounds can be prepared using milling or homogenization. Milling is achieved by combining a suspension of the substance with milling beads in a container and grinding the substance between the beads by rotating the container [13].

2.2. Proteins and ligands

Both proteins used in this project are human proteins expressed in *Escherichia coli*.
Protein A is a functional trimer of 129 kDa consisting of 354 amino acids. It is a metalloenzyme with Mn^{2+} as a cofactor.

Protein B has several isoforms with a size of 14-18 kDa depending on isoform. It is a disulphide rich protein with Ca^{2+} as a catalytic cofactor.

For project A, 7 different ligands were used; 2271, 9692, 6585, 1997, 2349, 9872 and 8246. The number combinations are based on internal identification codes for the different molecules. All ligands had confirmed specific binding to the protein and all had been tested with X-ray crystallography before in attempts to form complex structures with protein A. For protein B, two isoforms were used combined with one ligand, 0661. Information was gathered for the different ligands using an internal database. Also, standard assays were ordered internally regarding measurements of logD, solubility and pKa values.

IC50 values were available for the ligand 0661 and were for protein B1 0.91 µM and for protein B2 0.050 µM. In table 1 are listed further properties of the ligands used. Most values are determined through internal standard assays. Exceptions are logD calc. values which are calculated, some pKa values for which accurate literature values were available and the net charges which are calculated from the given pKa values as described in section 2.1.1. Neither

solubility nor logD could be obtained for all ligands but for one of them a logD value received from an earlier measurement using another method could be found, indicated by a star.

Table 1. Properties of the ligands. * means that it is an older measurement made with another method.

Ligand	Molecular weight (g/Mol)	Solubility aq. pH 7.4 (µM)	logD pH 7.4	logD, pH 7.4 calc.	pKa, A1	pKa, A2	pKa, B1	pKa, B2	Net charge pH 7.4
2271	174.99				2.6	10.3	10		0.00
9692	145.16	555	0.9	0.98			6.995		0.28
6585	421.46	482	3.2*	2.68	3.976				-1.00
1997	222.31	670	-0.5	1.49			10.89		1.00
2349	144.18	214	1.4	1.46			7.5		0.56
9872	148.17	306	-1	0.14			7.35	4.127	0.47
8246	146.19			-1.87	2.2		9.0	10.0	0.97
0661	446.46	<1	3		9.47				-0.01

2.3. Protein crystallization and structure determination

Many substances can transform into an ordered, crystalline state. In a protein crystal, individual molecules can adopt only one or a few orientations and create an ordered structure by non-covalent bonds. A crystal is built up of many identical unit cells of which each cell include all unique components repeated throughout the crystal [14].

When a crystal is hit by waves diffraction can occur, which is the interference created when waves hit an object with dimensions comparable to the wavelength. This interference can be constructive, if the interacting waves are in phase, or destructive, if the waves are out of phase. A diffraction pattern is created based on intensities which are increased for constructive interference and decreased for destructive. [15].

Each diffracted beam can be seen as coming from a set of parallel planes in the crystalline lattice, see figure 5. Each of these set of planes is given lattice indices, hkl, that defines the number of planes in the set per unit cell in the x, y and z directions, respectively.

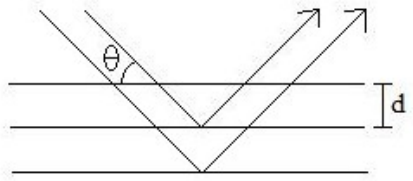


Figure 5. A simplified picture showing diffracted beams from planes in a lattice. d is the distance between the planes and θ the angle between the beam and the plane.

A set of planes with interplanar spacing d, gives constructive interference between incoming and outgoing beam, and thereby diffraction, if the angle θ with which the X-ray beam hits the lattice fulfils:

$$2d_{hkl}\sin\theta = n\lambda \qquad (2)$$

Equation 2 is called Bragg's law in which λ is the wavelength of the radiation and n is an integer [14].

In X-ray crystallography, X-ray radiation is used to obtain diffraction patterns of protein crystals in order to determine the structure of the protein. There are different types of X-ray sources, the most common ones used for protein crystallography being rotating anodes and particle storage rings. The most powerful is particle storage rings producing synchrotron radiation [14]. Synchrotron radiation is more intense and focused and often results in better resolution as well as shorter exposure time.

A reflection created from one diffracted X-ray can be described by a structure factor. The structure factor, F_{hkl}, can be written as a Fourier series with one term from each atom contributing to the reflection. F_{hkl} is a periodic function and thereby has amplitude, frequency and phase. The amplitude is obtained from the reflection intensity and the frequency from the X-ray source but the phase is unknown. This is called the phase problem and can be resolved in different ways. The method that has been used in this project is molecular replacement where a similar or identical protein with known phases is used as a model for calculating the phases of the protein in question. Other examples of methods include heavy-atom derivatives (e.g. Hg, Au, Se), for example isomorphous replacement and single wavelength anomalous dispersion. [14].

2.3.1. Growing protein crystals

Crystals of proteins are grown by precipitation from aqueous solutions with the help of precipitants. There are a lot of different precipitants, including salts, organic solvents and polymers (e.g. PEG) and combinations of different types of precipitants can be used. Other parameters also influence crystal growth, such as precipitant concentration, pH and temperature. In order to find suitable conditions for crystallization of a protein, screening is first performed over a range of different conditions to see when crystals are formed and after that optimization of those conditions is usually needed to improve the crystals [16].

The procedure of growing crystals usually involves addition of the precipitants to a water solution of the protein in a concentration just below what is needed for precipitation. After that, the water can evaporate slowly until precipitating conditions are reached which can then be maintained [14].

The most common method used for crystallization is vapor diffusion. In this method a droplet of protein solution mixed with a crystallization solution is deposited onto a cover glass. The crystallization solution often consists of buffer, salt and precipitant. The cover is placed onto a reservoir of crystallization solution and the difference in concentration between the drop and the reservoir will drive the system towards equilibrium. This will result in supersaturation of the protein solution and crystals will start to form when the system is at or close to equilibrium. For a salt solution, equilibrium will be met faster compared to a PEG solution [16].

Vapor diffusion can be performed in different ways, for example hanging drop and sitting drop, see figure 6. The shape of the drop is important and can affect the number of nucleation sites and thereby the crystal size [16].

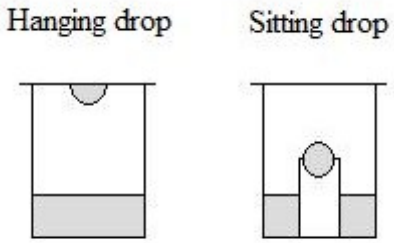


Figure 6. Different vapor diffusion setups; hanging drop and sitting drop. Adapted from Unge [16].

Fragile crystals might have to be stabilized against damage resulting from ligand introduction or freezing, described below. In this case, crosslinking could be an alternative. Lusty [17] describes one method of performing cross-linking using glutaraldehyde which was used for one of the proteins in this project. Potential problems with this method are a loss of diffraction quality and that it is difficult to control [17].

2.3.2. Ligand introduction

Protein-ligand interactions can be studied with X-ray crystallography. For this, crystals of protein-ligand complexes are needed. There are different ways of achieving this, including soaking and cocrystallization [3, 14].

Soaking: The method of soaking involves moving the protein crystals into crystallization solution that contains the ligand. The ligand can diffuse through water channels in the crystal and thereby reach the active site. Crystals made with this method are likely to resemble those of the protein without ligand [3, 14]. An advantage with this method is that crystals can be prepared and stored and soaked in different substances and conditions. It is a fast and convenient process and is easy to reproduce. A requirement for this method is that the protein crystals are functional for binding ligands. This can vary for different ligands and the soaking process can be affected by solubility, size and shape of the ligand [3].

Cocrystallization: In cocrystallization the ligand and protein are mixed and crystallized together [3, 14]. This method is more time-consuming and requires more protein. The crystal structure can differ with different ligands and might not be the same as for crystals of the protein without ligand and in some cases a new screening might be needed in order to achieve crystals. It can be good to use cocrystallization to confirm results obtained from a soaking experiment [3].

Some properties affect the binding of the ligand to the protein, including affinity and concentration as well as the solubility of the ligand. Even though values of solubility in water might give an estimate, it is important to know that precipitants are present which can affect the solubility and also that solvents, like dimethyl sulfoxide (DMSO), can be added to improve the solubility. The ionization state of functional groups might affect both solubility and binding which means that the pH of the solutions might affect the result [3, 18].

2.3.3. Data collection

Collection of X-ray crystallography data is often performed at temperatures of around 100 K, a technique called cryocrystallography. This is a way of reducing radiation damage which generally results in higher quality diffraction data with higher resolution [19].

The protein crystal, together with a drop of crystallization solution, is fished out with a nylon loop and quickly frozen in liquid nitrogen [14]. When doing this, there are risks regarding formation of ice crystals. To prevent this, the crystal is immersed in a cryoprotectant solution, often crystallization solution together with a cryoprotectant (e.g. ethylene glycol or PEG), before frozen. Ideally, enough concentration of a cryoprotective agent is present in the crystallization solution and the crystal can be frozen directly, reducing the amount of handling of the crystal [19].

A picture showing the setup for data collection is presented in figure 7.

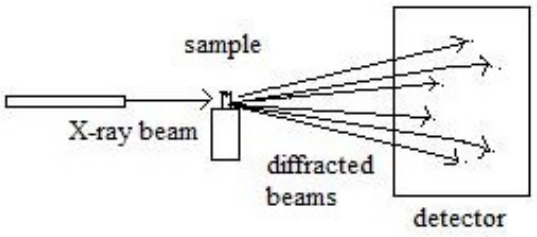


Figure 7. Simplified picture showing the experimental setup for X-ray crystallography.

The loop is mounted onto a goniometer head which rotates the crystal. The crystal is hit by an X-ray beam which after the sample is blocked by a beam stop to avoid extra radiation to enter the detector. There are different types of detectors, for example charge-coupled devices (CCD) and image plates [14].

In order to collect a complete dataset the crystal is rotated collecting diffraction patterns for each angle. The rotation range, i.e. how many degrees to collect, is determined by the symmetry of the crystal. A diffraction pattern, see figure 8, is called a reciprocal lattice because of the inverse proportionality between the distances between unit cells in the crystal lattice and corresponding distances on the diffraction pattern [14].

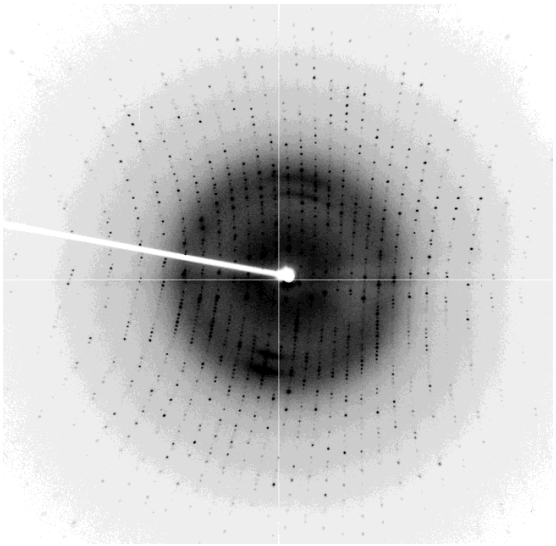


Figure 8. Example of a diffraction pattern.

The resolution of the data is the inverse of the distance from the origin for which reflections are obtained. The diffraction pattern has low resolution reflections in the centre and high resolution reflections close to the edges. Because of this, the resolution is dependent on the detector distance as can be seen in figure 7. For structure determination, protein crystals having a resolution of 3 Å or less are preferable [14].

Crystals are mosaics of several subcrystals and a reflection can be collected in different angles. The mosaicity is greater for protein crystals than other molecules since they are composed of flexible molecules that are held together by weak forces. Because of this, Measurements have to be performed over a small range of angles instead of just at one single angle [14].

2.3.6. Data processing

The result received from a data collection is a number of intensities, each with a set of indices. Processing is performed of the data including integration and scaling of intensities of identical reflections collected on different frames. The merging R factor, or Rsym, can be used to describe the agreement between different sets of data after scaling [14].

After the phase problem has been solved, by for example molecular replacement, the electron density, ρ(x,y,z), can be calculated from the structure factors. The model is placed into the electron density and after that structure refinement is performed in order to fit the model to the observed data using known stereochemistry of proteins. The R and Rfree factors are used to describe the model quality improvement during refinement. The Rfree is determined based on a random set of intensities that are set aside and not used during refinement [14].

2.4. Nuclear magnetic resonance spectroscopy

Magnetic resonance is the absorption of radiation by protons or unpaired electrons in a magnetic field. It can be applied on any nucleus with a non-zero spin but the most common is to use ^{1}H [15]. The spins of hydrogen nuclei can have two different orientations, either aligned with the field resulting in lower energy or aligned against the field with a higher energy [14]. Without a magnetic field, the net nuclear magnetic moment, or magnetization, is zero. Resonance can lead to transition of the spins into higher or lower energy states but for resonance to occur the radiation frequency must equal the frequency corresponding to the energy separation between the two states. The resonance condition that must be fulfilled is shown in equation 3:

$$\nu = \frac{\gamma B_0}{2\pi} \tag{3}$$

In equation 3, ν is the Larmor frequency, γ the magnetogyric ratio of the nucleus and B_0 the magnetic field. [15].

In nuclear magnetic resonance (NMR) spectroscopy a magnetic field is applied to the sample, leading to polarization of the spin orientation and a net magnetization. [15]. By applying a radiofrequency (RF) pulse at the Larmor frequency perpendicular to the field the magnetization is tilted into the xy plane. The magnetization starts to precess around the static magnetic field and gives rise to an oscillating signal in the RF coil [14]. The resulting signal from a pulse is called free-induction decay (FID) which is a time-domain signal where the oscillating, decreasing RF signal is plotted against time. The FID signal can be frequency analyzed into a frequency-domain spectrum which is called a 1D NMR spectrum [14]. The

integrations of the different signals, calculated as the area under the absorption lines, is proportional to the number of spins and can help to decide which chemical group contributes to each signal [15].

Magnetic moments in nuclei interact with a local magnetic field. The local field depends on the applied magnetic field but can differ due to an induced electronic current in the molecule by the field. How large this contribution is depends on the electronic structure next to the nucleus, expressed by the shielding constant, σ, of a nucleus. This varies for each atom and affects the resonance frequency, or Larmor frequency [15]:

$$\nu_L = (1-\sigma)\frac{\gamma B_0}{2\pi} \tag{4}$$

It is common to talk about the chemical shift, δ, of a nucleus which is based on the difference on the resonance frequency of the nucleus in question, ν, and that of a reference, ν_0, where:

$$\delta = \frac{\nu - \nu^o}{\nu^o} \times 10^6 \tag{5}$$

The net magnetization vectors for nuclei with different chemical shifts precess at different frequencies meaning that the FID contains characteristic absorption frequencies from which the chemical shifts can be determined [14].

The spin system will, after a while, return to equilibrium by a mechanism called spin-lattice relaxation, where spins loose energy to the surroundings. The spin-lattice relaxation time constant is denoted T_1. A spin-spin relaxation also occurs depending on the spreading of the phases of identical spins. This happens because of the exchange of energy between spins due to coupling. The spin-spin relaxation time constant is denoted T_2 [14].

Different nuclei can interact with each other through spin-spin coupling. Coupling between two nuclei results in splitting of the absorption signal into two lines which are separated by a distance called the coupling constant, J. For coupling to occur the nuclei cannot be more than a few bonds apart, meaning that the splitting of signals can help decide which groups are neighbors [14].

2.4.1. Concentration determination

One application of NMR, which has been used in this project, is to measure concentrations of substances. Since the area of a peak is proportional to the number of protons that peak corresponds to, relative concentrations can be determined. To be able to do this, a standard is needed with a structure different from that of the studied substance resulting in a sharp single peak a distance from the peaks from the molecule of interest. A good standard is important for a precise result; an example is Trimethylsilyl propionic acid (TMSP), see figure 9 [20].

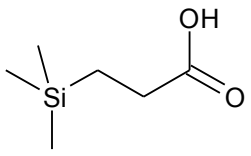


Figure 9. Molecular structure of TMSP.

The precision of the concentration measurement have been determined to 1% for concentrations above 20 mM [21]. For lower concentrations, long experiment times are needed since more scans are required for an adequate signal to noise ratio. However, measurements of concentrations are possible for concentrations of 1 mM with a precision of 5%. The precision is affected by for example the choice of integrals and integral tails. If possible, single sharp peaks should be chosen [20].

The concentration of the substance can be determined from the spectrum by integrating the peaks from the standard as well as the substance. The concentration is calculated using equation 6 where c and c_{ref} denotes concentration of substance and standard, respectively, A and A_{ref} the peak areas and n and n_{ref} the number of protons the peaks correspond to. The peak resulting from the standard TMSP, used in this project, is situated at approximately 0 ppm and corresponds to nine protons.

$$c = \frac{A}{A_{ref}} * \frac{n_{ref}}{n} * c_{ref} \tag{6}$$

2.4.2. Dissociation constants

Another application of NMR used in this project is the possibility to measure protein-ligand binding giving values of the dissociation constant, K_D, described by equation (1).

$$K_D = \frac{[P][L]}{[PL]} \tag{7}$$

In the equation above, [P] is the concentration of free protein, [L] the concentration of free ligand and [PL] the concentration of complex [22].

Dissociation constants can, in several ways, be determined by the help of NMR. One way is by titrating ligand into a protein solution until excess concentration and monitor the NMR signal. In the competition binding experiment, used in this project, this is done in the presence of a reporter ligand which binds to the same active site as the studied ligand. During the NMR experiments the signal of the reporter is studied [22].

If the dissociation constant of the reporter ligand, L_1, is known, K_D for the studied ligand, L_2, can be determined using NMR experiments and equations derived in Appendix A. This results in equation 8.

$$K_{D,2} = \frac{\alpha}{P_0 - [PL_1] - \alpha}\left(L_2 + [PL_1] + \alpha - P_0\right) \tag{8}$$

Where $\alpha = \frac{[PL_1]K_{D,1}}{L_1 - [PL_1]}$ (9)

In the equations above, P_0 is the total protein concentration, L_2 the total concentration of the studied ligand, L_1 the total concentration of the reporter ligand and $[PL_1]$ the concentration of reporter ligand bound to protein. $[PL_1]$ is unknown and can be found from the NMR

experiments. If the NMR signal from the reporter ligand is studied, the maximum height, I_{max}, is received without protein present and the minimum signal, I_{min} is received when protein has been added. When a competing ligand, L_2, is added the signal from L_1 regains in proportion to how the population $[PL_1]$ decreases resulting in signal intensity I. The regain in signal can be described by equation 10:

$$F = \frac{(I_{\max} - I)}{(I_{\max} - I_{\min})} = \frac{[PL_1]}{[PL_1]^0} \tag{10}$$

Where $[PL_1]^0$ is the concentration of reporter ligand bound to protein without a second ligand present. This value can be derived from equation 7 as:

$$[PL_1]^0 = \frac{1}{2}(L_1 + P_0 + K_{D,1}) - \left[\frac{1}{2}(L_1 + P_0 + K_1)^2 - P_0 L_1\right]^{1/2} \tag{11}$$

3. Materials and Methods

This section will cover the performance of the project. The project included protein crystallization, soaking with ligands and structure determination by X-ray crystallography to study protein-ligand complex formation. NMR spectroscopy was used for solubility studies and affinity measurements.

3.1. Dissolution of ligands

The ligands were each mixed with five different solvents or solvent mixtures; deuturated DMSO (Deutero GmbH, 99.8%), H_2O (milliQ water), hydroxypropyl-β-cyclodextrin (Roquette) 28% (w/w) in water (prepared by Kalle Sigfridsson), a mixture of DMA (Alfa Aesar, 99%), PEG400 (Hampton Research) and water (DMA:PEG400:H_2O 1:1:1) and HCl (Hampton Research) or NaOH (Hampton Research, >98%) depending on the ligand being an electron donor or acceptor. HCl or NaOH were mixed to a molar ratio of 1:1 with the ligands.

For project A, 7 different ligands were used; 2271, 9692, 6585, 1997, 2349, 9872 and 8246, all with confirmed specific binding to the protein. Three of these ligands, 2271, 9692 and 8246 had successfully formed complexes with the protein in earlier structure experiments. The procedure for dissolving the substances in each solvent was started by mixing the substance with the solvent to a concentration of 500 mM and vortexing the sample. If not dissolved, heating for 5 minutes at 50°C were performed on a heating plate. If still not dissolved, sonication for 5 minutes were performed. If not dissolved at this stage, more solvent was added giving a concentration of 400 mM. This was repeated, for every 100 mM decrease in concentration, until the substance was dissolved or the concentration was down to 200 mM.

For project B one ligand was used, 0661, which was known to have low solubility but relatively high affinity to the protein. Because of the low solubility, it was decided to mix it to a concentration of 100 mM. The same solvents were used except HCl/NaOH. Additionally, in this project, two types of nanosuspensions of the ligand in water were prepared and used for soaking. The different nanosuspensions used were an amorphous and a crystalline preparation of the ligand with substance concentrations of 10 mM and they were obtained from Kalle Sigfridsson. The preparation methods are described by Sigfridsson et al. [13].

3.2. Crystallization

The crystallization part was started by growing protein crystals for project A, using already established crystallization conditions communicated by Rob Horsefield. For project B, already grown crystals were provided by Cristian Bodin. Once large enough to harvest, the formed protein crystals were soaked with the different combinations of ligands and solvents.

3.2.1. Crystallization of protein A

Precipitant solutions were mixed consisting of 0.35 M ammonium citrate tribasic, pH 7 (Hampton Research, >97%) and 14-19% PEG3350 (Hampton Research). 24 well plates were filled up with 500 µl of precipitant solution in each well with different concentrations of PEG3350 in each column. Protein A were used that were dissolved to 10 mg/ml in a buffer consisting of 50mM Hepes (4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid), 100mM KCl, 100µM $MnCl_2$, 2mM TCEP (tris(2-carboxyethyl)phosphine) and 5% Glycerol pH 8.1. The hanging drop method was used for crystallization where 1 µl of protein together with an equal amount of precipitant solution was deposited onto a cover which was screwed onto each well. The crystals were left to grow for at least two weeks in 20°C.

3.2.2. Soaking

Soaking solutions were prepared for the different proteins. For protein A, a stock of ammonium citrate pH 7.4 was first prepared by dissolving ammonium citrate (tri ammonium citrate anhydrous, Fluka, ≥98%) in H_2O and setting pH by adding NaOH and HCl until reaching pH 7.4 while stirring with a magnetic stirrer. The choice of pH was based on the conditions used for the NMR binding experiments. H_2O was added until a concentration of 1 M was reached. For the soaking solution, this ammonium citrate solution was mixed with PEG3350 and H_2O giving 0.35 M ammonium citrate and 17% PEG3350. For reference with ligand 9692 a second soaking solution at pH 8.4 was prepared by mixing 0.24 M lithium citrate pH 8.4 with 17% PEG3350 and water. The reason for this was to mimic the conditions used in the earlier successful soaking experiment with this ligand.

For protein B, two isoforms of the protein were used for which prepared crystals were obtained from Cristian Bodin. The soaking conditions used were based on the crystallization conditions. For protein B1, the soaking solution prepared consisted of 45% PEG400 and 100 mM Bis-Tris, pH 6 (Hampton Research) and for protein B2, a soaking solution was prepared using 3.5 M sodium formate (Hampton Research) together with 100 mM Hepes, pH 7.5 (4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid sodium salt, Hampton Research).

For project A, soaking experiments were performed in 24 well plates filling a well with 500 µl soaking solution. A soaking drop was prepared on a cover containing 9 µl soaking solution mixed with 1 µl dissolved substance giving a concentration of substance 1/10 that of the dissolved concentration, resulting in 20-50 mM. For substances that had not dissolved in the lowest concentration tested, a small substance particle was added in addition to the liquid to make sure the ligand was present in the soaking drop.

For protein B1, the soaking experiments were performed as for project A, resulting in 10 mM concentration of substance, except for nanosuspensions. For the nanosuspensions, two soaking experiments were performed for each preparation, since the concentration of substance in the nanosuspensions was only 10 mM. One was done as before resulting in a concentration of substance of 1 mM. In the other experiments, the soaking drop was prepared by exchanging the water in the soaking solution by suspension to maximize the concentration of ligand in the drop, resulting in a concentration of 4.5 mM. For B2, soaking experiments were only performed with DMSO, cyclodextrin and the crystalline nanosuspension depending on results achieved for B1. Soaking with the nanosuspension was done with the second method reaching a ligand concentration of 4 mM.

To each soaking drop, about 5-10 crystals were added and the cover was then put over the soaking solution reservoir. For protein B2, an additional crosslinking step was performed before soaking. 5 µl of glutaraldehyde solution (25% in H_2O, Sigma Aldrich) was put on a microbridge and 100 µl of precipitant solution in the well and the cover glass with a drop containing protein crystals was placed above. After 30 min incubation the crosslinking experiment was stopped and the crystals were washed in precipitant solution before transferred to the soaking drops.

For project A, experiments with two references, 2271 and 9692, were done to test the experimental setup. After that, an experimental design was set up for the following ligands taking into account different temperatures and solvents in order to limit the number of experiments. The experiments performed are shown in table 2.

Table 2. Experimental plan over soaking experiments were X marks an experiment that was performed. The first two ligands were references that were done before the rest of the experiments were planned and are not part of the experimental design.

Solvent	T (°C)	2271	9692	6585	1997	2349	9872	8246
H_2O	20	X				X	X	
	30	X	X	X	X			X
DMSO	20	X	X	X		X	X	
	30		X		X			X
Cyclodextrin	20	X			X			X
	30		X	X		X	X	
DMA:PEG:H_2O	20	X	X	X	X			X
	30		X			X	X	
HCl/NaOH	20			X		X		
	30		X		X		X	X

The plates were incubated in the different temperatures, 20°C and 30°C, according to the experimental design during 48 h except for the reference ligand 2271 for which 24 h was known to be enough.

In project B, only 20°C were used depending on results achieved with project A and soaking was done overnight.

3.2.3. Co-crystallization

Apart from soaking experiments, ligand 6585 was also co-crystallized together with protein A because of the large size of the ligand. The ligand was dissolved in DMSO to 500 mM. 1 µl of that solution was mixed with 81 µl of the protein solution giving a ligand concentration of 6 mM and a protein concentration of 9.9 mg/ml. The mixture was incubated for approximately one hour on ice and an additional couple of hours in room temperature.

In order to find a different crystal form more prone to accommodate the ligand an initial screen was performed. Two standard screens were used, Jena Bioscience JBScreen Classic HTS I and JBScreen Classic HTS II, were an automated setup were used using a Mosquito (TTP LabTech) creating drops containing 200 nl protein-ligand solution mixed with 200 nl stock solution. A minor screen was also performed by hand around the condition successful for the protein with PEG3350 concentration of 10-20% combined with ammonium citrate concentrations of 0.2-0.3 M, pH 7, with drops of 1+1 µl.

Optimization steps were performed after evaluating the result from this screening. Two plates were done by combining PEG3350 concentrations of 18-29% with ammonium citrate, pH 7, concentrations of 0.2-0.35 M. Another plate was prepared combining 20-30% PEG4000 with 0.1-0.2 M ammonium sulfate (Hampton Research) and 0.1 M tri sodium citrate (Sodium citrate tribasic dehydrate, Hampton Research) at pH 5.6 and 6.

3.3. X-ray crystallography

X-ray crystallographic data were collected from frozen crystals and the data was processed in order to determine if complexes had formed.

3.3.1. Freezing and testing of crystals

After soaking, the crystals were fished with nylon loops (Hampton Research) and directly frozen in liquid nitrogen. The loops were put in vials which were loaded into SPINE standard

pucks. Between 3 and 10 crystals were frozen from each condition. The crystals were screened in-house using one of three equipments; FR-E+ rotating anode generator (Rigaku) and Saturn A200 CCD detector (Rigaku) connected to an Actor robot (Rigaku), the latter remade to accommodate SPINE standard pucks, FR-E+ generator and R-axis HTC image plate detector (Rigaku) or Rigaku FR-E generator (Rigaku) and R-axis HTC detector connected to a modified Actor robot. Two frames at 1 and 91° were taken. The crystal with the best combination of diffraction quality and resolution for each condition was chosen for data collection. Data collection was performed collecting images in 0.5° steps between 1 and 150° for project A and between 1 and 120° for project B. Most data sets were collected on the equipment described first, with the differing data collections marked in table D1, Appendix D.

Some crystals were sent to ESRF (European Synchrotron Radiation Facility) for data collection if a high enough resolution of 2.7 Å was not obtained in-house. This data was collected on beam line ID23 1 with an ADSC (Area Detector Systems Corporation) detector.

3.3.2. Data processing

An automated script, developed by Hans-Georg Beisel, was used for data processing in project A. The script used autoPROC [23] or mosflm [24] for indexing and XDS [25] or mosflm to integrate the data and Scala [26] for scaling. A reference Rfree set was used for which 5% of the reflections were excluded from refinement. Further, molecular replacement was done using PHASER [26]. Here, the monomer of a previously in-house solved crystal structure was used as a starting model to find three copies in the asymmetric unit. Finally, refinement was performed with autobuster [27] and automatic rebuilding with RAPPER [26].

In some cases, if the script did not work or the parameters indicated too low quality, processing was done using another automated script. The processing was then performed by using imosflm [24] manually for indexing and integration and a script for further processing including scaling with Scala, molecular replacement with molrep [26] or PHASER and refinement with autobuster or refmac [26, 28]. 5% of the reflections were excluded for Rfree. The resolution cutoff was determined by looking at completeness, higher than 95% overall, and mean I/σ, higher than 2.0 in the outer shell.

For project B, all data was treated with the latter method described above.

The created maps were examined in coot [29]. Some manual refinement of the protein and addition of water molecules and metal ions was performed. A few cycles of autobuster were run and the ligands were fitted manually into the resulting difference density. If, after additional autobuster refinement, there was electron density covering the ligand at a σ level of 1, it was decided that a complex had formed. None of the data sets in project A were refined completely since the complexes that were obtained were already known and completely refined earlier.

3.4. Concentration determination

The concentration determination experiments started by preparing samples which were then incubated before the concentrations were determined using NMR spectroscopy. The conditions used for the experiments were chosen to mimic the soaking conditions as much as possible.

3.4.1. Preparation of samples

For project A, samples for concentration measurements with NMR spectroscopy were prepared for each combination of ligand and solvent. Dissolved substance was added to soaking solution, pH 7.4, described in section 3.2.2. The concentration of substance used in the samples was the same as in the soaking experiment, which was 1/10 that of the dissolved concentration. TMSP (2,2,3,3-D(4)-3-(trimethylsilyl) propionic acid sodium salt, Cambridge Laboratories) was used as an internal standard and was dissolved in H_2O to a concentration of 100 mM. TMSP was added to the NMR sample to a concentration 1/10 that of the ligand.

40-50 µl of each sample was transferred to a 1.7 mm NMR capillary using a Hamilton syringe, 10 µl (Sigma-Aldrich). Different syringes were used for different solvents and they were washed between each sample with acetone and H_2O. Duplicate tubes were filled with each sample. The tubes were sealed with stearine and incubated in 20°C and 30°C, respectively, for 48 h to mimic the soaking conditions. The tubes put in 30°C were weighed before and after incubation to detect possible evaporation.

For project B, the samples were prepared in the same manner, using the different soaking conditions, except for nanosuspensions which were not included. Incubation was done overnight and only in 20°C based on the results received from project A, except for the ligand dissolved in DMSO where both temperatures were used as a control.

3.4.2. Running and processing of samples

The NMR experiments were performed manually on a 600 MHz NMR spectrometer (Oxford AS600). A sample with 1 mM glucose in H_2O/D_2O was used to lock and adjust the field. The samples were run without lock since an addition of heavy water would have altered the conditions compared to the soaking experiments. The samples were shimmed using gradient shimming selective on PEG giving a sharper and narrower peak than water.

For project A, one pulse experiments were performed with a recycling time, d1, of 20 s. The number of scans, ns, was adjusted depending on the predicted concentration of substance to receive high enough signal to noise ratio. For 50 mM ns=16, for 40 mM ns=32, for 30 mM ns=48 and for 20 mM ns=64. The NMR experiments were run with a single 90° pulse and presaturation of water were used to improve the baseline since the water peak from the experiments were broad due to interactions with other components in the samples.

For project B, ns was set to 72 due to the low concentrations. To keep the experiment time down, d1 was set to 15. For the experiments with protein B1, where the solution did not contain PEG, the water peak was used for shimming.

The spectra obtained were phased manually and baseline corrected around the peaks of interest. The TMSP peak was used as a reference and by comparing the integrations of the TMSP peak and chosen peaks from the ligand the concentration of ligand in the sample could be calculated according to equation 6.

Some experiments were repeated in order to determine an error margin. In these cases new samples were prepared and run in the same manner as before. Mean values and standard deviations were calculated for these experiments. Further, pooled standard deviations were calculated for each temperature and in total as described in appendix B. Confidence intervals were calculated from the total pooled variance using a t-distribution and 95% confidence level, both for the difference between two mean values and for a single value.

3.5. Determination of dissociation constants

NMR was used to determine the dissociation constants (K_D) for the different ligands used for project A. From a plate containing all ligands dissolved in DMSO to 100 mM, dilutions were made with d6-DMSO to concentrations of 10 mM, and for two ligands even to 1 mM. Measurements to validate the concentrations were performed using a buffer containing 50 mM Tris-d6, 20 μM $MnCl_2$ and 10% D_2O. The ligand 8246 was not included in the experiments due to its insolubility in DMSO.

The same buffer mixed with the reference substance, 100 mM, was used for K_D determination experiments. NMR experiments were run according to the T2 filter experiments described by Dalvit [30], while observing the reporter. A first round of NMR was run with the buffer and 2 μM of protein before adding ligand in 7 rounds giving higher and higher concentrations leading to displacement of the reporter. A protocol of the rounds can be found in Appendix A. A final round was performed adding the ligand 2271, known to be a strong binder. Some experiments with blank samples, only adding DMSO, were also performed for control. For two ligands, 2271 and 6585 another set of experiments was performed with lower concentrations since they were expected to yield lower K_D values.

The received spectra from each round were integrated on the peak from the reporter and the different values plotted together to visualize the change in signal. These integrals could be used to determine K_D values for some of the ligands according to equation 8.

3.6. Correlations of data

The TIBCO spotfire software (version 3.1.0, TIBCO Software Inc.) was used to plot all data received for project A in different combinations in order to more easily find and illustrate correlating data. PCA analyses were performed with the software Simca P+ (version 12.0.1.0, Umetrics) for the concentration measurements to find which ligands and solvents that gave similar results.

4. Results

The results achieved from the different parts of the project will be presented as tables and plots with additional comments. In the end of this section correlations of data from the different parts as well as the properties listed in section 2.2. will be presented mainly as scatter plots.

4.1. Dissolution of ligands in solvents

For project A, the different ligands were dissolved in the solvents as described in section 3.1. In table 3, the concentrations at which each ligand dissolved, or if it did not dissolve at all, are listed as well as whether heat or sonication was needed.

Table 3. Resulting concentrations at which each ligand dissolved in each solvent. H and S regards whether heat or sonication, respectively, was used to dissolve the ligand at this concentration. * means that the ligand was not dissolved, [a] that HCl was used and [b] that NaOH was used.

Solvent		2271	9692	6585	1997	2349	9872	8246
H_2O	Conc. (mM)	500	200*	200*	200*	200*	500	500
	H/S	-	H+S	H+S	H+S	H+S	-	-
DMSO	Conc. (mM)	500	500	500	500	500	500	200*
	H/S	-	-	-	-	-	H	H+S
Cyclodextrin	Conc. (mM)	500	200*	200*	200	200*	500	500
	H/S	H	H+S	H+S	H+S	H+S	-	-
DMA:PEG:H_2O	Conc. (mM)	500	200	300	300	500	500	500
	H/S	-	H	H	H	-	H	H
HCl/NaOH	Conc. (mM)		500[a]	500[b]	200[a]*	500[a]	500[a]	500[a]
	H/S		-	H+S	H+S	-	-	-

As can be seen in table 3, most ligands dissolve in DMSO as well as HCl/NaOH in high concentrations. H_2O and cyclodextrin gives almost identical results, in many cases the ligand has failed to dissolve. Heating was successfully used in many cases while sonication in a few cases helped the ligand to dissolve.

For project B, a concentration of 100 mM was used to dissolve the ligand. DMSO was the only solvent used that dissolved the ligand at this concentration.

4.2. Concentration determination

The concentration in the soaking condition was measured with NMR. In Figure 10 can be seen an example of an obtained spectrum.

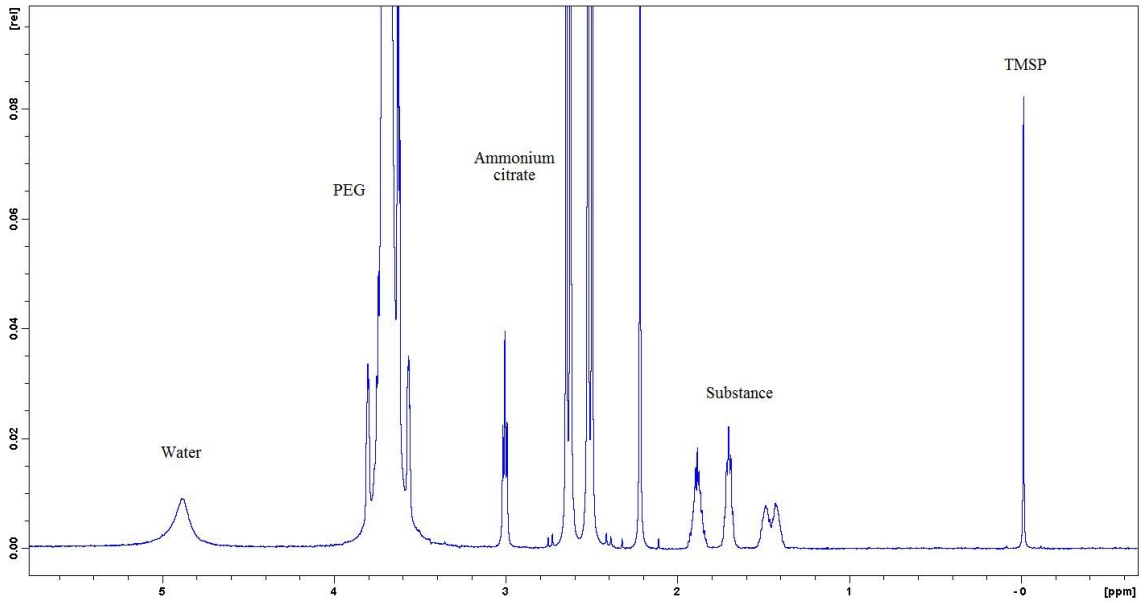


Figure 10. NMR spectra for substance 8246 dissolved in DMSO. Highlighted are peaks from water, PEG3350 and ammonium citrate present in the soaking condition and peaks from the standard TMSP as well as the tested substance.

To the left in the spectrum are highlighted the remains of the suppressed water peak, in the middle peaks from the soaking condition of PEG3350 and ammonium citrate and to the right some of the peaks from the substance as well as the peak from the standard TMSP. Weighing of the samples incubated in 30°C showed a mean loss of mass of 0.2% which was regarded as negligible.

The resulting concentrations determined from the integrated peaks from the experiments with project A can be seen in table 4. In cases of repeated experiments, which are listed in Appendix E, a mean value is shown as an underlined number. The numbers represent the concentration in mM that was calculated from the spectra. The numbers in brackets regards the concentration added to the sample, meaning the maximum theoretical concentration.

Table 4. Concentrations in mM measured by NMR of the different ligands in the different soaking conditions. Values in brackets are the maximum possible concentrations depending on added amount of substance. * denotes cases where the ligand was not dissolved in the solvent and the added amount of substance is unknown. Underlined values represent mean values received from repeated experiments.

Solvent	T (°C)	2271	9692	6585	1997	2349	9872	8246
H_2O	20	57 (50)	0 (20*)	0 (20*)	22 (20*)	5 (20*)	43 (50)	45 (50)
	30	42 (50)	0 (20*)	0 (20*)	23 (20*)	5 (20*)	44 (50)	45 (50)
DMSO	20	56 (50)	36 (50)	4 (50)	48 (50)	45 (50)	48 (50)	8 (20*)
	30	45 (50)	36 (50)	9 (50)	45 (50)	47 (50)	48 (50)	8 (20*)
Cyclodextrin	20	50 (50)	15 (20*)	1 (20*)	23 (20)	24 (20*)	62 (50)	59 (50)
	30	45 (50)	15 (20*)	1 (20*)	30 (20)	24 (20*)	61 (50)	49 (50)
DMA:PEG:H_2O	20	44 (50)	20 (20)	2 (30*)	33 (30)	44 (50)	38 (50)	62 (50)
	30	45 (50)	24 (20)	5 (30*)	30 (30)	41 (50)	39 (50)	65 (50)
HCl/NaOH	20		49 (50)	2 (50)	7 (20*)	46 (50)	47 (50)	40 (50)
	30		49 (50)	3 (50)	6 (20*)	46 (50)	48 (50)	41 (50)

It can be seen from table 4 that in most cases the solubility in the soaking condition is close to, or at, the upper limit of what was added, the main exceptions being 6585 and cases where the ligand did not dissolve in the solvent.

A table of the experiments that were repeated can be seen in Appendix E. A sample mean and sample standard deviation was calculated for each set of measurements and the resulting values can be seen in the same table. The pooled standard deviations calculated out of these experiments resulted in 5.6 for 20°C, 7.0 for 30°C and 6.4 over both temperatures. The confidence interval for the difference of two means, μ_1-μ_2, each a mean from two data points resulted in a confidence estimate of $\bar{x}_1 - \bar{x}_2 \pm 13$. Likewise, the confidence interval for a single value also has an upper and lower confidence limit of 13.

In figure 11, the solubility has been plotted against ligand where temperature and solvent are marked by shape and color, respectively.

Figure 11. Measured solubility (mM) against ligand where the different solvents are marked with different colors and temperatures with different shapes.

In figure 11, the difference in solubility between temperatures can be seen as well as between ligands. It can be noted the small variation in solubility between temperatures for each combination. It can also be seen that there is a large difference between ligands both how soluble the ligands are in the soaking solution and which cosolvent that gives the best and worst result.

In figures 12 and 13 are found PCA diagrams comparing different ligands as well as solvents.

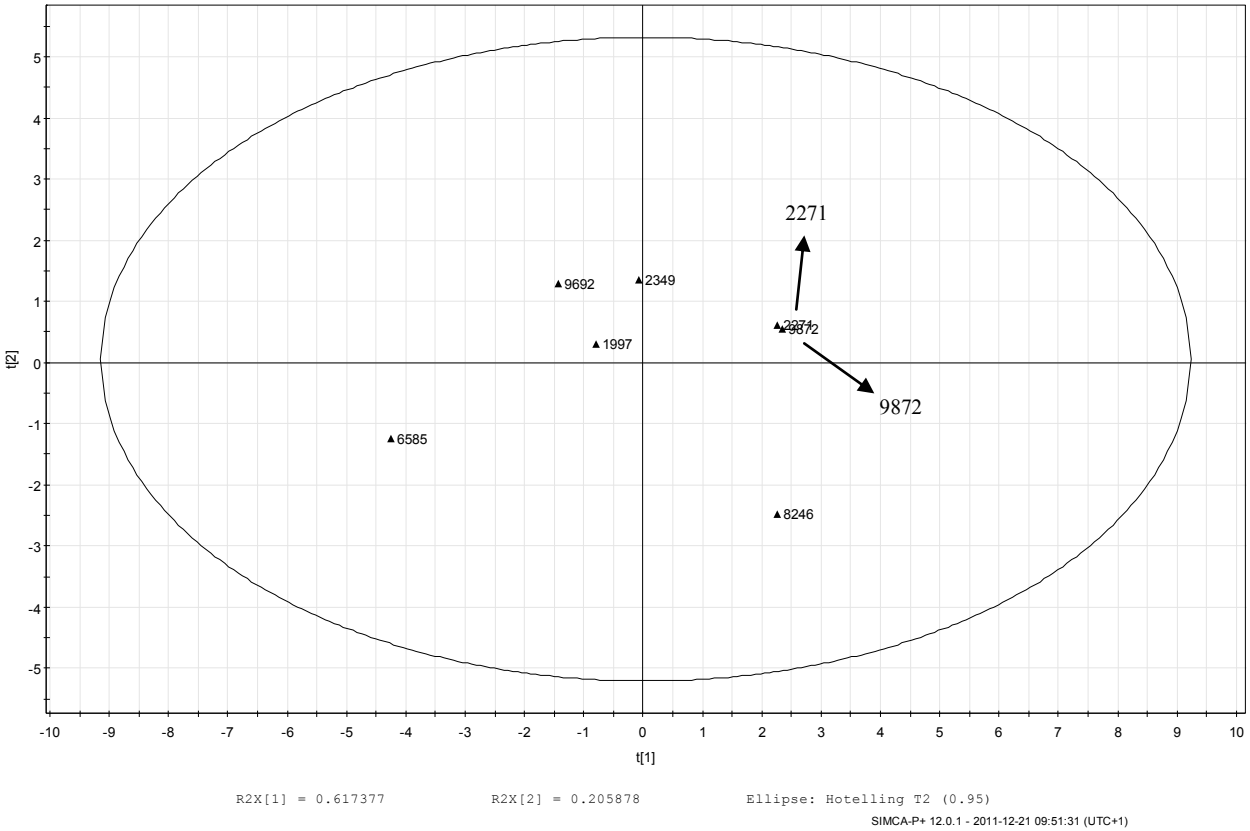


Figure 12. PCA diagram highlighting similarities and differences in solubility between ligands.

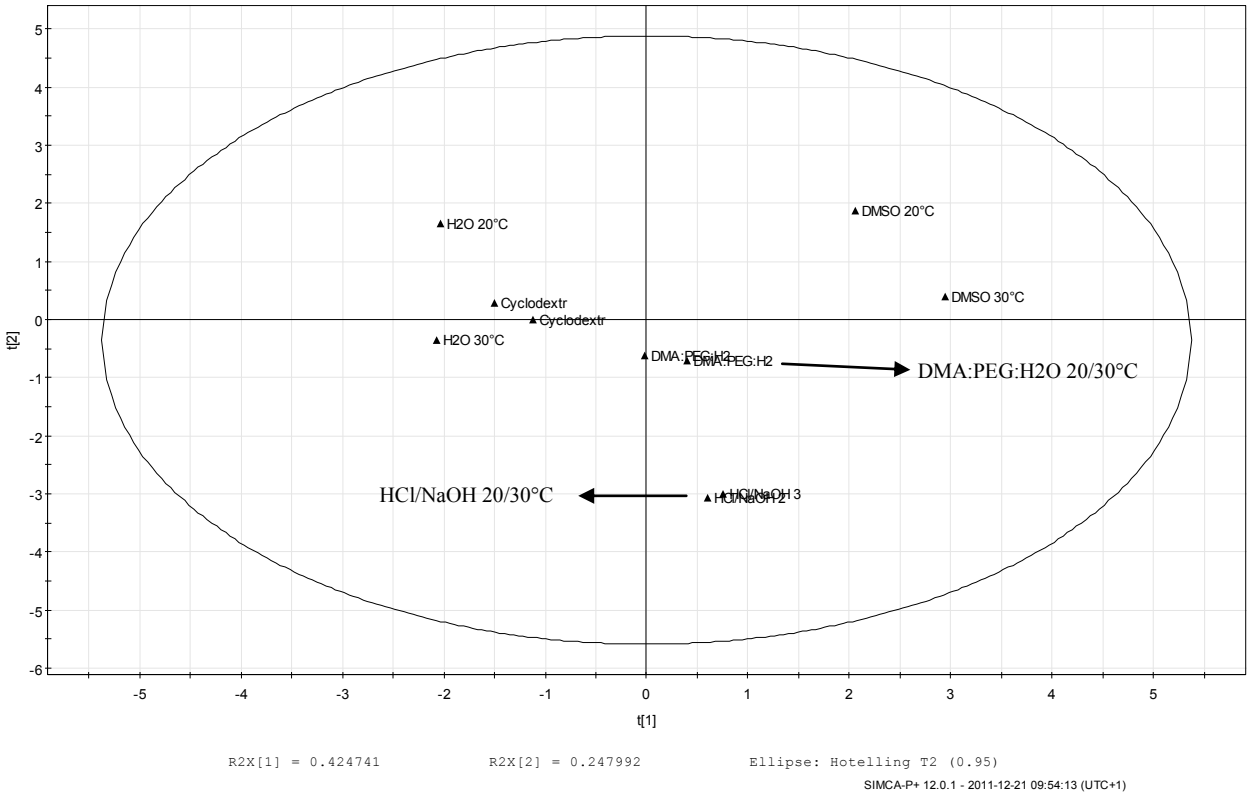


Figure 13. PCA diagram highlighting similarities and differences in solubility between solvents.

The ligands most similar to each other regarding solubility based on these analyses are 2271 and 9872 and the most similar solvents water and cyclodextrin.

The concentration determination experiments for project B2 resulted in no visible peaks from the substance in any of the cosolvents. For B1, in some cases peaks could be detected but very small peaks that were difficult to integrate. In two cases a concentration could be determined, however with some uncertainty regarding the accuracy. These were for DMSO 0.5 mM and for cyclodextrin 0.2 mM, of the possible 10 mM.

4.3. Crystallization and structure determination

Crystallization of protein A resulted in formed crystals within five days although at this time very small. They were left to grow at least two weeks when the crystals were large enough to harvest. Even though crystals of varying size and shape were obtained the relatively large number of crystals made it possible to choose large crystals with a uniform shape. Examples of grown crystals of protein A is shown in figure 14.

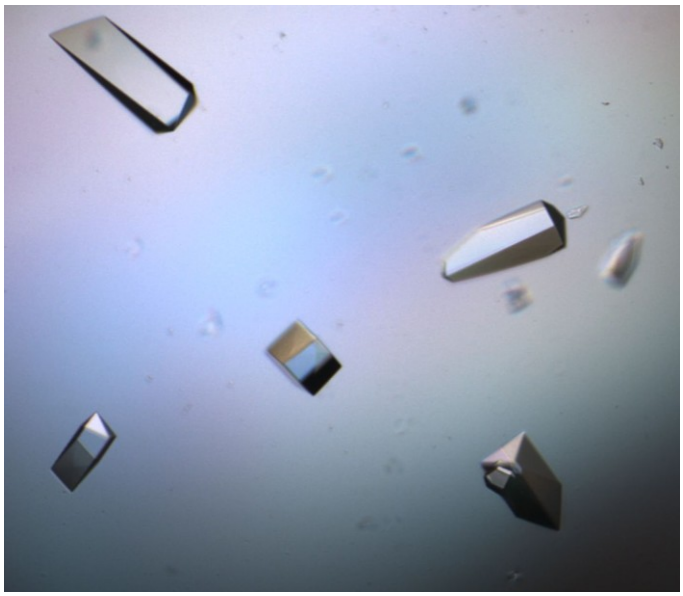


Figure 14. Example of protein A crystals.

During soaking, some of the crystals cracked but left enough smaller pieces to freeze. The crystals survived freezing without cryoprotection and data were collected. The diffraction patterns varied but an example can be seen in figure 15.

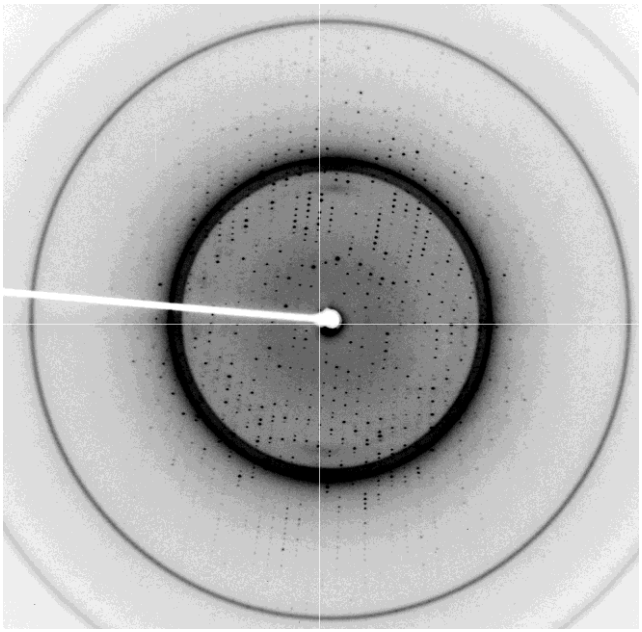


Figure 15. Diffraction pattern resulting from data collection of 2349/HCl.

In figure 15 can be seen significant solvent rings which were generally present for protein A crystals. Data collection of protein A also often resulted in multiple diffraction patterns, high mosaicity and anisotropy. The diffraction patterns from protein A crystals generally resulted in resolutions around 2.5 Å in house.

The data collection of protein B1 crystals resulted in diffraction patterns like the one seen in figure 16.

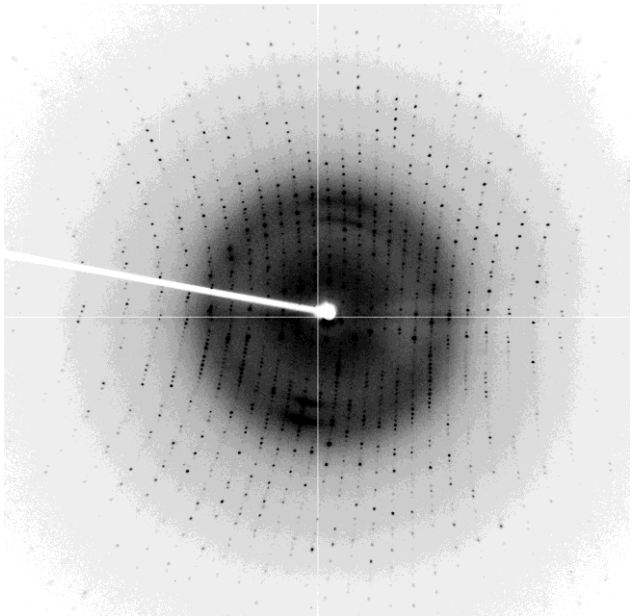


Figure 16. Diffraction pattern resulting from data collection of 0661/B1/DMSO.

Comparing the diffraction of protein A and B1 it can be noticed in the latter the lack of the solvent rings seen in figure 15. Protein B1 crystals also gave higher resolutions, less than 2 Å. Crystals of protein B2 showed weaker diffraction than both protein A and B1 and only in one case a high enough resolution for data collection in house was obtained.

The data received from the data collections were processed as described in section 3.3.2. In Appendix B crystallographic and refinement data for the different data sets for both projects can be found. Most structures were solved and the cutoffs defined in section 3.3.2 were met.

The results regarding complex formation for project A are shown in table 5. The results are given as the number of subunits of the protein, out of three possible, in which protein-ligand complexes had formed. For two data sets, the R/Rfree values, which can be found in appendix D, indicated that the structure was not solved and they were excluded from the results.

In section 3.3.2. is described the method used to decide whether a complex had formed or not. After looking through all data, it was sometimes very clear that a complex had formed and sometimes very clear that it had not. However, after looking at 40 datasets, it was clear that some of the data in between was more close to being a complex than others, were some density could be observed even if it was not possible to refine. If a more generous interpretation regarding when a complex has formed is used the numbers in brackets are achieved.

Table 5. Result from the X-ray crystallography experiments for project A. The numbers indicate in how many subunits, out of three, complex formation has occurred. The numbers in brackets are the result of a wider interpretation of the data. – regards data sets were the structure was not solved.

Solvent	T (°C)	2271	9692	6585	1997	2349	9872	8246
H_2O	20	2 (3)				0	0 (1)	
	30	2	0	0	0			2
DMSO	20	3	0	0		0	0 (1)	
	30		0 (1)		0			1 (2)
Cyclodextrin	20	2 (3)			0			2 (3)
	30		0	0		0	0	
DMA:PEG:H_2O	20	2	0 (1)	-	-			2 (3)
	30		0			0 (1)	0	
HCl/NaOH	20			0		0 (1)		
	30		0		0		0 (1)	3

It can be seen in table 5 that two ligands, 2271 and 8246, form complexes with the protein in all conditions. For 9692, 2349 and 9872 complex formation occurred in some cases using the more generous interpretation. However, the ligands could not be unambiguously modeled into the density.

A soaking experiment with the ligand 9692 was also performed at pH 8.4 to mimic earlier successful experiments. Complex formation did not occur when using the first, stricter definition but a positive result was obtained with the second interpretation method.

The cocrystallization experiment with 6585 resulted in crystals diffracting well enough for data collection at the synchrotron. No structural changes of the protein could be seen and no complex had formed.

For protein B1, the ligand formed complexes with the protein in both of the two subunits when dissolved in DMSO. This data was fully refined and deposited in an internal structure

database. In the other cases, complexes were formed in one subunit. For protein B2, the ligand formed complex with the protein in both subunits when dissolved in DMSO. For none of the other soaking experiments with protein B2 a high enough resolution for data collection in house was received and no synchrotron beam time was available to give results to include in this report.

4.4. Dissociation constants

Dissociation constants, K_D, for project A were determined during the project as described in section 3.5. The resulting values are listed in table 6.

Table 6. Measured K_D values for the ligands used in project A. * means that the value was calculated as a mean from earlier measurements.

Ligand	2271	9692	6585	1997	2349	9872	8246
K_D, protein A (µM)	0.76*	5010	130	7940		1260	

For 2271, K_D had been measured earlier in house and was calculated as a weighted mean out of four earlier measurements, obtained from Per-Olof Eriksson. The dissociation constant for 8246 could not be measured in this experimental setup because of its insolubility in DMSO. 2349 were not possible to receive data for as the signals did not differ significantly from the effect of the vehicle.

4.5. Correlations of data

In figure 17 the number of formed complexes is plotted against the measured solubility for project A with ligands marked with color. In the upper plot the number of complexes regards the first interpretation of the results. In the lower plot, the more generous interpretation is used, described in section 4.2, resulting in the number of complexes here called complexes 2.

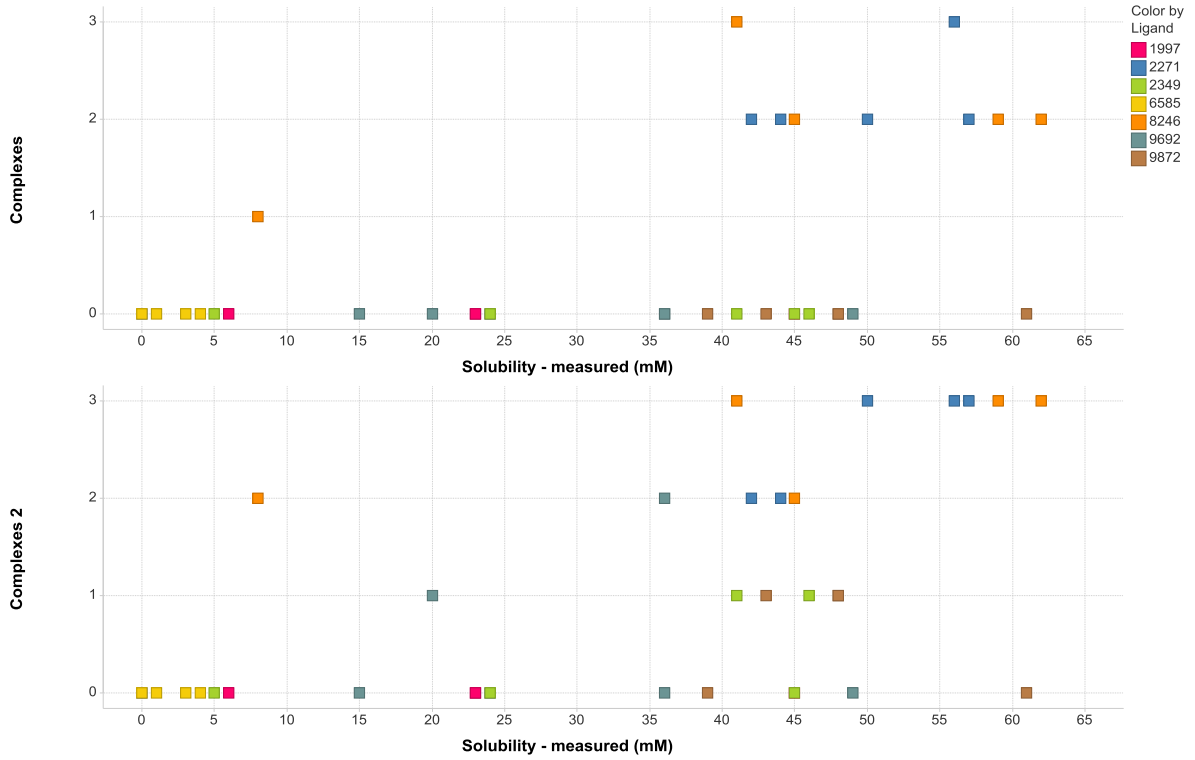


Figure 17. The number of formed complexes against solubility (mM) for project A, using strict and generous interpretation, respectively.

In figure 17 can be seen that, in these experiments, more complexes have formed at higher solubilities with a few exceptions.

In figure 18 measured solubilities against measured and calculated logD values for project A are plotted.

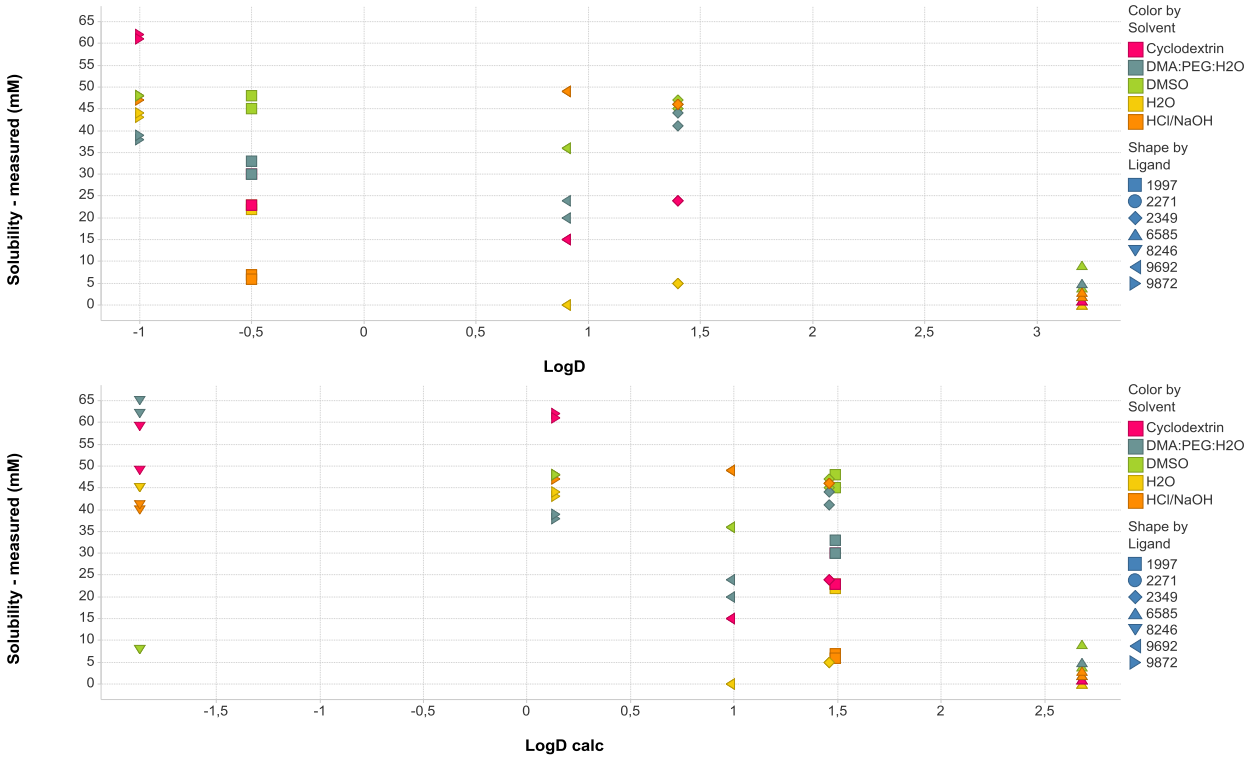


Figure 18. Solubility (mM) against logD and logD calc, respectively.

From the lower graph in figure 18 can be seen that substances with logD values over zero have higher solubility in DMSO compared to water and cyclodextrin while the opposite is true for the ligands with a negative logD. DMA:PEG:H_2O is generally in between. In the upper graph, on the other hand, all substances have a higher solubility in DMSO.

In figure 19 are plotted the number of formed complexes against the dissociation constant for project A, with the different interpretation methods regarding the number of complexes.

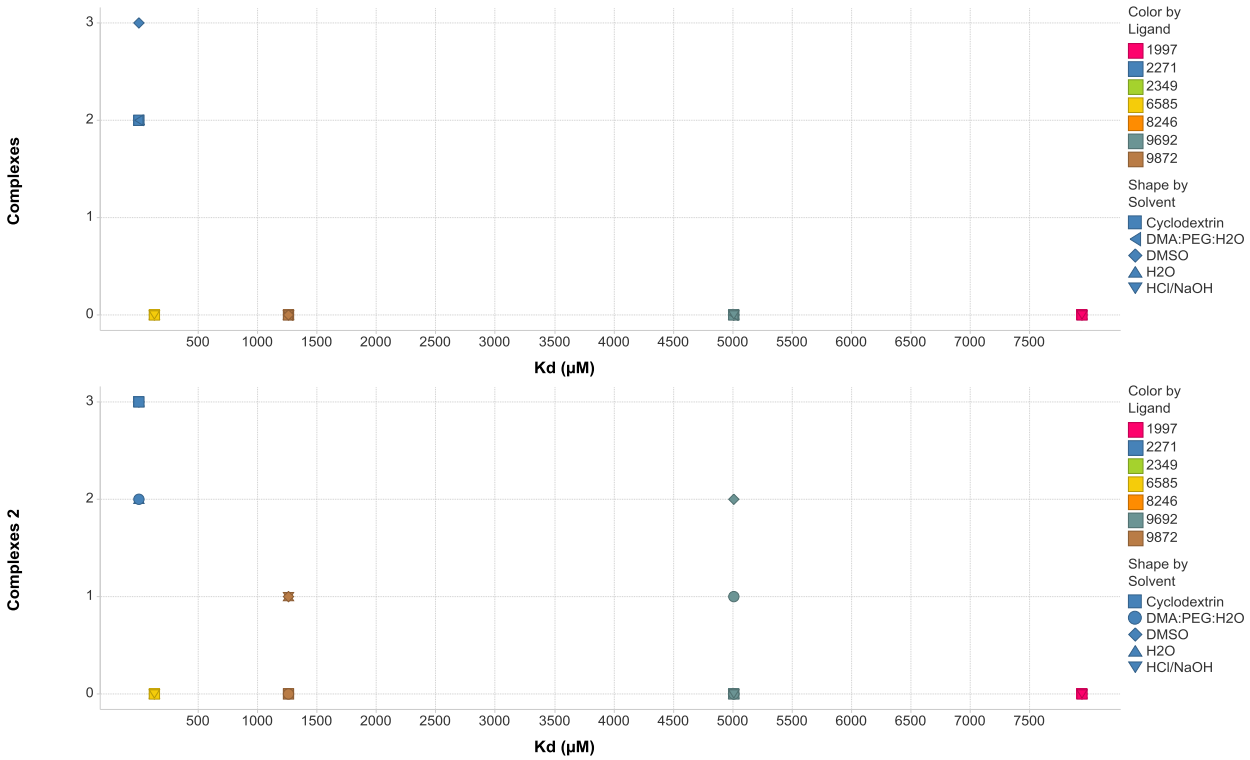


Figure 19. Number of formed complexes against K_D (µM) for project A, using strict and generous interpretation, respectively.

It can be seen that 2271, having the lowest of the determined dissociation constants, are the only ligand resulting in complexes with the strict interpretation method.

For project B can be noted the ligand's low solubility and highly positive logD from table 1. This can be connected to the fact that in most cases the solubility in the soaking condition was too low to be detected but gave the highest concentration in DMSO. Low IC50 values of 0661 to protein B1 of 0.91 µM and to protein B2 of 0.050 µM can be correlated to the fact that complexes were formed in all conditions for which data were collected.

5. Discussion

Interpretations of the results obtained will be discussed here as well as correlations between the different results in order to draw conclusions regarding future utilization.

5.1. Solubility

The project showed that the solubility experiments performed with NMR were possible to make receiving good quality data. This is a promising result since similar experiments measuring solubilities in soaking conditions are very unusual. There are several things that complicate the measurements. Thin capillaries, more complicated to handle, had to be used in order to receive high enough concentration of ligand without using a lot of substance which could be difficult to obtain. Also, including heavy water in the sample is preferable when using NMR spectroscopy and this was not done in this project since the soaking conditions should be mimicked. However, the magnet showed high stability and small drift since a high resolution was received without using lock resulting in sharp signals. The water peak was quite wide despite the high amount of water, probably due to interactions with other components. Integration and calculation gave reasonable values and some were repeated to give an idea about the reproducibility and error margin. By repeating the experiments from the start variations coming from sample preparation, NMR measurements and integration could be included. Confidence intervals with limits of 13 were received for a single value as well as for the difference between two means. This was seen as acceptable for the aim of this project. If this kind of experiments will be done again further repeats would be recommended to give a more exact result with less uncertainty.

There are some uncertainties regarding the solubility measurements. In cases when the ligand did not dissolve in the solvent, the amount of ligand transferred to the soaking solution is unknown meaning that the highest possible concentration in the soaking solution is uncertain. Comparing solvents and ligands, these measurements do not regard the possibility that the ligand dissolve better in the soaking condition than in the solvent. This is possible due to the fact that PEG is included in the soaking condition which is a good solvent [11]. This could also be seen in an earlier master thesis work where a PEG condition resulted in increased solubility for small, poorly soluble compounds compared to a salt condition [31]. More accurate and comparable results could have been achieved if more than the dissolved substance had been added giving the same amount of added substance in all samples and resulting in a saturated mixture. Aggregation of substance would have made the experiments more difficult to achieve practically, something that was already noticed as a problem for the poorly soluble compounds.

Regarding the confidence limits obtained when looking at table 4 and figure 11 it can be concluded that based on these data no difference in solubility between temperatures 20 and 30°C can be detected. Because of this, only one temperature was used for experiments with project B. Even though it might differ between conditions, the experiments in this project suggest that raising the temperature with 10°C will not improve solubility in the soaking solution for 48 hours incubation. Other temperatures could be used for new experiments, giving a wider range, but that could be more practically difficult to perform.

Comparing tables 3 and 4 only a small difference can be seen between the solubility in the cosolvent and the solubility in the soaking solution. There are some exceptions, like 6585 as well as 0661 in project B which were both highly soluble in for example DMSO but precipitated in the soaking solution. These results suggest that a high solubility in the solvent

does not automatically mean a high solubility in the soaking solution. However, if comparing different solvents, the one with the highest solubilizing properties for the ligand is likely to give the highest solubility even in the soaking solution, which can be seen for example with ligand 0661 which showed best results for DMSO.

From table 3 can be seen that the dissolution of the ligands in project A resulted in generally high concentrations in DMSO. The ligand for project B did only dissolve in DMSO. Since DMSO is the most commonly used solvent for this purpose and is generally known to be a good solvent this is not surprising. Regarding the solubility in the soaking condition, from table 4, it can be seen a trend that DMSO in many cases gives the opposite results than water which is logical, since they are very different types of solvents.

It can be seen from table 3 that dissolving the ligands in HCl/NaOH gives concentrations either higher than or similar to water. This is expected since ionization of compounds often leads to higher solubility. It can be seen that HCl/NaOH in two cases, 9692 and 2349, when the solubility in water is low increases the solubility significantly also in the soaking condition.

As can be seen in figure 13, water and cyclodextrin gives similar results regarding solubility which can also be seen on the dissolved concentrations in table 3. It is hard to say exactly why this is but one explanation could be that the ligands in some cases, for some reason, do not form complexes with cyclodextrin. This is discussed by Del Valle, as mentioned in section 2.1.2., that the complexation could be affected if the substance is poorly soluble [8]. It could be interesting to try different solvents for cyclodextrin and also to try other concentrations of cyclodextrin in the solution but it did not fit in the scope of this project. The ligands that dissolve well in water might bind to cyclodextrin but from these results it is not possible to tell if that is the case. If higher concentrations would have been tried for the ligands that dissolve well a difference might have been detected that could indicate that complexes had formed between ligands and cyclodextrin.

The mixture DMA:PEG:H_2O was hoped to be able to dissolve many substances since it is a mixture of different types of solvents. It fulfills its expectations since it seldom gives the lowest concentration even though it not often gives the highest either. Instead it generally is in between water and DMSO, based on the results in table 4 and figure 11. The strength of this solvent system is the possibility to give an adequate solubility of several different kinds of ligands, making it a good option in crystallization. This use is limited, however, by the fact that DMA is quite toxic and has to be handled in the fume hood. A possibility could be to replace DMA with DMSO but it is uncertain if the same effect would be obtained.

Heating seems to help dissolving the ligands in most cases but sonication had a very limited effect. Based on these results, sonication would not be recommended for this purpose in this kind of experiments. Heating was performed at only 50°C for 5 min to limit the risk of evaporation and the risk of degradation of ligands.

Based on the results achieved there can be seen a difference in solubility depending on the cosolvent used. It differs between ligands, however, so there is not one solvent that is generally better. In figure 12 can be seen that two ligands show very similar results; 2271 and 9872, which both showed high solubility in all solvents. These ligands are very different structurally which shows that it is difficult to correlate the solubility to the structure of the ligand.

Comparing the results obtained for 0661 to the values in table 3, it can be seen that this ligand is much more difficult to dissolve compared to all ligands used for project A. The lack of accurate concentration data in the soaking condition makes it hard to draw significant conclusions regarding difference in solubility between different conditions. However, the results that B1 gives measurable results compared to B2 suggests that the soaking condition for B1, including PEG, gives an increased solubility compared o the salt condition for B2, which agrees with earlier discussions. The measured solubility results for project B is quite uncertain and similar values might have been achieved for other solvents but in some cases the peaks from the solvent interfered with the peaks from the ligand complicating the integration. It is, however, clear that the compound has very poor solubility.

5.3. Crystallization and structure determination

The crystals obtained seemed to be more sensitive in some conditions than others but no solvent or ligand did systematically affect the crystals more than any other. The X-ray crystallography experiments were performed without using a cryoprotection step. The reason for this choice was mainly the risk of washing away ligand and good enough results were believed to be achieved without it. Some experiments were performed with cryoprotection as well without significantly improved data quality. An alternative to eliminate the risk of washing away ligand could have been to include substance in the cryoprotectant solution but that would have increased the workload as well as the ligand consumption. There is also a possibility that the cryoprotectant would have affected the solubility or compete with the ligand in binding to the protein.

For two of the data sets, the structures were not solved. However, this is not believed to have affected the results and conclusions significantly, especially since these ligands were not close to forming complexes with the protein in any of the other solvents. For all obtained data, more processing and refinement could further have improved the quality but was not believed to be needed for this project.

The placement of the ligands into the weak electron density might have been facilitated, had the fragments contained electron rich atoms. This is especially the case for small, symmetrical fragments, including several ligands in project A. It was somewhat difficult to make an interpretation and draw a line between what was decided to be a complex and not. Because of this, two interpretation methods were used as described in section 4.3.

The studies with project A did not result in complexes with ligands that had not earlier been shown to form complex with the protein if using the first, strict, interpretation method. This is not very surprising since most of the ligands tested are small fragments with high dissociation constants. That the ligand 9692 that had formed complex in earlier experiment performed by others did not bind to the protein, despite attempts of reproducing the experiment and discover any differences in experimental setup, is hard to explain. It could show a lack of reproducibility but could also be a coincidence or some difference in interpretation of results since there in several of the data sets collected for 9692 were almost enough electron density to fit the ligand. It might be interesting to repeat the experiment additional times to investigate if a positive result can be achieved. If the ligand is on the boundary of being a strong enough binder to form complex with the protein the small differences that occur between experiments due to protein batch, ligand preparation and soaking performance might make the difference between a formed complex or not.

For 6585, cocrystallization was tried. Since this ligand showed high affinity but was too large to fit in the binding site and showed low solubility, cocrystallization seemed like a good option. The formation of a complex structure could be dependent on crystal packing. However, initial screening resulted in crystals with the same space group as for the ligand-free protein, no structural differences could be detected and no complex was formed. More experiments could be tried using more conditions, longer incubation time or other ligand concentrations but it did not fit in the scope of this project.

As for solubility, the data obtained suggests that a difference in temperature of 10°C have no effect on complex formation. If also regarding the complexes achieved with the more generous interpretation method this conclusion is strengthened. Also here, it could be interesting to try other temperatures as well giving a larger difference. There is, however, uncertainties regarding the effect on the protein crystals which would require experiments to test the tolerance of the crystals.

Complexes were formed with all solvents. Even if this project mainly shows that DMSO is the best cosolvent to use the second interpretation method shows that for the ligand 2349, HCl and DMA:PEG:H_2O gave better results than DMSO which gives a small inclination that other solvents might be better and could be worth trying and also that the best solvent to use is ligand-dependent. It is also worth highlighting that all the cosolvents tried resulted in formed complexes and could be alternatives for cases where DMSO is not possible to use, for example if the crystals have low DMSO tolerance or if the solvent molecules inhibits the protein-ligand binding.

An earlier master thesis project at the department has investigated different cosolvents' effect on crystallization using cocrystallization [32]. Several solvents resulted in crystal formation and complex formation and it was concluded that by trying different cosolvents more ligands could probably be solubilized. It was also concluded that alternative cosolvents are interesting especially for projects with weak binders. This is in agreement with the results obtained from this project where the stronger binder 0661 seems to be less affected by a low solubility.

It is interesting that complexes formed using cyclodextrin. No information regarding the use of cyclodextrin earlier for crystallization has been found and it is an interesting alternative due to its differing properties compared to other solvents. When regarding the binding of the ligand to the protein, a prerequisite is that the ligand has not too high affinity to cyclodextrin compared to the protein which could prevent the release. Also worth regarding is the affinity of water to cyclodextrin since the water molecules could be used to compete with the ligand resulting in release of ligand from the hydrophobic cavity. It would be interesting to know how high these affinities are for the molecules used in this project. A hypothesis is that by saturating the surrounding water, a change in equilibrium when molecules bind to the protein crystal could cause the cyclodextrin to release more and more ligand molecules which would be interesting to investigate further. Even though complexes did form with cyclodextrin in this project there is a possibility, as discussed earlier, that inclusion complexes have not formed making it hard to draw conclusions regarding the process based on these results.

The ligand 0661 has formed complexes with protein B1 in all cases, even with nanosuspensions despite the lower concentration. The only data collection made for protein B2 is for the ligand dissolved in DMSO and that showed complex formation. This is especially interesting since experiments performed earlier with the same conditions did not

result in complex crystal structure which is one of the reasons for which this ligand was chosen, combined with the combination of low solubility and high affinity which suited the purpose of this project. These results show that complex formation is possible for all solvents tried and that using nanosuspensions of ligands can be an option in crystallographic experiments. However, nanosuspensions are more difficult to prepare compared to using an ordinary cosolvent, especially in high concentrations, and more compound is needed.

5.5. Correlations of data

Regarding the K_D determination, the values received have large errors because of few data points resulting from the measurements as well as the fact that the effect of the vehicle was not regarded. Still some trends can be seen since 2271 has a significantly lower K_D than the other ligands and 6585 and 9872 follows. The low signals from 2349 indicate that the K_D value is higher than the values received for the other ligands. Even though the K_D values are high, all ligands have been determined as specific binders to the protein previously in house, communicated by Per-Olof Eriksson, meaning that complex formation is still possible. From figure 19 can be seen that 2271 was the only ligand with a determined dissociation constant that formed complexes with the protein. Regarding both interpretation methods, it can be seen that 6585 with the second highest value still do not result in complex formation indicating that dissociation constant is not the only criteria that determines complex formation. The lack of positive data and the uncertainty of K_D values make it difficult to draw significant conclusions of correlations between dissociation constant and binding.

In figure 17, the number of complexes for project A is plotted against solubility. It shows a tendency that higher solubility increases the chance of forming complexes. It is strengthened when using the lower plot where the second interpretation method is used. Further experiments with other proteins and/or ligands resulting in more formed complexes would make it easier to draw definite conclusions. It can be seen, however, that high solubility is not enough to bind to the protein crystal. It can be discussed that high solubility together with a low dissociation constant could give a high chance of binding but it is hard to draw such a conclusion from these results.

The experiments with protein B shows that complex formation is possible even with ligands poorly soluble in the soaking solution. In this case, however, the ligand is a much stronger binder than the majority of the ligands used for protein A, shown in the low IC50 values. This would again suggest that the chance of forming complex depends on a balance between solubility and binding strength of the ligand. This is in agreement with discussions made by Podjarny et al. [3], regarding the importance of the relationship between K_D and the ligand concentration compared to the protein concentration on complex formation. This relationship causes the solubility of the ligand to often be limiting, especially for weak ligands.

For the values of logD and pKa, data is gathered from different sources. This could make the comparison less precise but all data are believed to be relatively accurate. It can, however, be seen that the measured and calculated values of logD differ significantly in some cases. It is difficult to say why this is and it is something that might be good to look into but it did not fit in the scope of this project. When regarding logD connected to solubility as seen in figure 18, some correlation can be seen when looking at the calculated values of logD. The results suggest that positive logD values gives higher solubilities in DMSO and lower in water and cyclodextrin and the opposite in cases of negative values. This fits well with the definition of logD as being the logarithm of the pH-dependent distribution of the substance in octanol compared to water [7]. DMA:PEG:H_2O being in between can also be connected to the logD

values since it is a mixture of different types of solvents. The plot with the measured logD values results in a higher solubility for DMSO even for negative values. Based on these results, logD values could be a help in selecting which cosolvent to use for different ligands but the accuracy of the value should be regarded.

The solubility was also plotted against net charge in hope of correlating the values to solubility in water and HCl/NaOH but no trend could be seen. It could be interesting to use different pH values in the future and investigate if an effect can be seen.

6. Conclusions

- The results from the project suggests that the choice of cosolvent affect the solubility in the soaking condition
- Knowing the logD value of the ligand can help in deciding which solvent that works best
- The temperature does not seem to affect either solubility or complex formation when changing from 20 to 30°C
- Complexes could be obtained for all cosolvents tested and also when using nanosuspensions
- Higher solubility in the soaking condition seems to increase the possibility to obtain complexes but more data would be needed to confirm this
- For ligands with high dissociation constant solubility in the soaking solution seems to have a larger impact on complex formation than for ligands with lower dissociation constant, based on these results
- For the future, similar experiments with more ligands, proteins and possibly solvents could be performed in order to obtain more data points and be able to confirm these results and draw more conclusions. Other parameters could also be included like pH.

References

[1] Erlanson D.A. McDowell R.S. O'Brien T. Fragment-based drug discovery. Journal of medicinal chemistry. 2004; 47(14): 3463-3482.

[2] Hartshorn M.J. Murray C.W. Cleasby A. Frederickson M. Tickle I.J. Jhoti H. Fragment-based lead discovery using X-ray crystallography. Journal of medicinal chemistry. 2005; 48(2): 403-413.

[3] Podjarny A., Dejaegere A., Kieffer B. (ed.). Biophysical Approaches Determining Ligand Binding to Biomolecular Targets: Detection, Measurement and Modelling. Cambridge: Royal Society of Chemistry; 2011.

[4] Balakin K. V., Savchuk N. P., Tetko I. V. In Silico Approaches to Prediction of Aqueous and DMSO Solubility of Drug-Like Compounds: Trends, Problems and Solutions. Current Medicinal Chemistry. 2006;13:223-241.

[5] Moore D. S. Amino Acid and Peptide Net Charges: A Simple Calculational Procedure. Biochemical education. 1985;13(1):10-11.

[6] Augustijns P, Brewster M. E. (ed). Solvent systems and their selection in pharmaceutics and biopharmaceutics. New York: Springer/AAPS Press, 2007.

[7] Xing L, Glen R. C. Novel Methods for the Prediction of logP, pKa, and logD. Journal of Chemical Information and Computer Sciences. 2002;42:796-805.

[8] Del Valle E. M. M. Cyclodextrins and their uses: a review. Process Biochemistry. 2004;39(9):1033-1046.

[9] Stella V. J., He Q. Cyclodextrins. Toxicologic Pathology. 2008;36:30-42.

[10] Sigfridsson K., Lundqvist A., Strimfors M., Svanberg P. Pharmacokinetics of BX211 after oral administration of nano- and microsuspensions to rats during early development. Asian Journal of Pharmaceutical Sciences. 2009;4 (3):161-168.

[11] Strickley R. G. Solubilizing Excipients in Oral and Injectable Formulations. Pharmaceutical Research. 2004;21(2):201-230.

[12] Barnes R. K., Viegas T. X., Harris J. M., Viegas T. X. Poly(ethylene glycol) in AccessScience, McGraw-Hill Companies [Internet]. 2008. [Cited 2011-12-13]. Available: http://www.accessscience.com.proxy.lib.chalmers.se.

[13] Sigfridsson K., Forssén S., Holländer P., Skantze U., de Verdier J. A formulation comparison, using a solution and different nanosuspensions of a poorly soluble compound.
European Journal of Pharmaceutics and Biopharmaceutics. 2007;67:540–547.

[14] Rhodes. Crystallography made crystal clear. 2nd ed. San Diego: Academic Press, 2000.

[15] Atkins P., de Paula J. Physical Chemistry. 8th ed. Oxford: Oxford University Press, 2006.

[16] Unge T. Crystallization methods. In: Bergfors T. M. (ed). Protein crystallization – techniques, strategies and tips, a laboratory manual. La jolla: International university line, 1999.

[17] Lusty C. J. A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. Journal of Applied Crystallography. 1999;32:106-112.

[18] Danley D. E. Crystallization to obtain protein–ligand complexes for structure-aided drug design. Acta Crystallographica. 2006;D62:569–575.

[19] Garman E. F. Crystallization for Cryo-Data Collection. In: Bergfors T. M. (ed). Protein crystallization – techniques, strategies and tips, a laboratory manual. La jolla: International university line, 1999.

[20] Rizzo V., Pinciroli V. Quantitative NMR in synthetic and combinatorial chemistry. Journal of Pharmaceutical and Biomedical Analysis. 2005;38:851–857.

[21] Maniara G., Rajamoorthi K., Rajan S., Stockton G. W. Method Performance and Validation for Quantitative Analysis by ^{1}H and ^{31}P NMR Spectroscopy. Applications to Analytical Standards and Agricultural Chemicals. Analytical Chemistry. 1998;70(23):4921–4928.

[22] Fielding L. NMR methods for the determination of protein–ligand dissociation constants. Progress in Nuclear Magnetic Resonance Spectroscopy. 2007;51:219–242.

[23] Vonrhein C., Flensburg C., Keller P., Sharff A., Smart O., Paciorek W., Womack T., Bricogne G. Data processing and analysis with the autoPROC toolbox. Acta Crystallographica. 2011;D67:293-302.

[24] Leslie A. G. W., Powell H. R. Processing Diffraction data with Mosflm. In: Read R.J., Sussman J. L. (ed). Evolving Methods for Macromolecular Crystallography. Springer;2007. pp. 41-53.

[25] Kabsch W. Xds. Acta Crystallographica. 2010;D66:125-132.

[26] Collaborative Computational Project, number 4. The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallographica. 1994;D50:760-763.

[27] Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P., Sharff A., Smart O., Vonrhein C., Womack T. Cambridge, United Kingdom: Global Phasing Limited. 2010.

[28] Murshudov G. N., Skubák P., Lebedev A. A., Pannu N. S., Steiner R. A., Nicholls R. A., Winn M. D., Longa F., Vagina A. A. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica. 2011;D67:355–367.

[29] Emsley P., Lohkamp B., Scott W., Cowtan K. Features and Development of Coot. Acta Crystallographica. 2010;D66:486-501.

[30] Dalvit C. Theoretical Analysis of the Competition Ligand-Based NMR Experiments and Selected Applications to Fragment Screening and Binding Constant Measurements. Concepts in Magnetic Resonance. 2008;A32(5):341–372.

[31] Johansson C. Investigation of the effect of agents used in protein crystallisation on ligand solubility [Master of Science Thesis]. Göteborg: Chalmers University of Technology. 2006.

[32] Bengtsson A. Cosolvent Impact on Protein Crystallisation [Master of Science Thesis]. Göteborg: Chalmers University of Technology. 2005.

Appendix A: K_D determination from competitive experiments

Two ligands, L_1 and L_2, are in equilibrium with one binding site on a protein P. The dissociation constants are defined as follows where L_1 is the reporter and L_2 the test ligand.

$$K_{D,1} = \frac{[P][L_1]}{[PL_1]} \tag{A1}$$

$$K_{D,2} = \frac{[P][L_2]}{[PL_2]} \tag{A2}$$

In equation A1, [P] is the concentration of free protein, $[L_1]$ the concentration of free reporter ligand, $[PL_1]$ the concentration of ligand bound to protein and similar for the test ligand in equation A2.

The total protein concentration $P_0 = [P] + [PL_1] + [PL_2]$
The total reporter ligand concentration $L_1 = [PL_1] + [L_1]$
The total test ligand concentration $L_2 = [PL_2] + [L_2]$

Using this information, two equations can be derived:

$$[PL_1]K_{D,1} = (P_0 - [PL_1] - [PL_2])(L_1 - [PL_1]) \tag{A3}$$
$$[PL_2]K_{D,2} = (P_0 - [PL_2] - [PL_1])(L_2 - [PL_2]) \tag{A4}$$

If $K_{D,1}$ is known there are three unknown variables left; $K_{D,2}$ which is the searched value, $[PL_1]$ and $[PL_2]$. $[PL_1]$ can be found from the NMR experiment as described below.

If L_2 were not present the equilibrium for L_1 would be:

$$PL_1 \leftrightarrow P + L_1 \text{ where } K_{D,1} = \frac{[P][L_1]}{[PL_1]} \tag{A5}$$

From equation A5, the concentration of reporter ligand bound to protein without a second ligand present, $[PL_1]^0$, can be derived as:

$$[PL_1]^0 = \frac{1}{2}(L_1 + P_0 + K_{D,1}) - \left[\frac{1}{2}(L_1 + P_0 + K_1)^2 - P_0 L_1\right]^{1/2} \tag{A6}$$

During the NMR experiments, if the NMR signal from the reporter ligand is studied, the maximum height, I_{max} is received without protein and the minimum signal, I_{min} is received when protein has been added. When a competing ligand, L_2, is added the signal from L_1 regains in proportion to how the population $[PL_1]$ decreases resulting in signal intensity I. The regain in signal can be described by equation A7:

$$F = \frac{(I_{max} - I)}{(I_{max} - I_{min})} = \frac{[PL_1]}{[PL_1]^0} \quad \text{(A7)}$$

Combining equations A3 and A4, $K_{D,2}$ can be determined resulting in equation A8:

$$K_{D,2} = \frac{\alpha}{P_0 - [PL_1] - \alpha}(L_2 + [PL_1] + \alpha - P_0) \quad \text{(A8)}$$

Where $\alpha = \frac{[PL_1]K_{D,1}}{L_1 - [PL_1]}$ (A9)

$[PL_1]$ is obtained from experiments using equation A6 and A7.

Appendix B: Variance and confidence intervals

A population has a mean value, μ, and a standard deviation, σ, the latter being a measure of the variance of data. For a set of data a sample standard deviation, s, can be calculated according to equation B1 where $\bar{x}$ is the sample mean and n the sample size [B1].

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \tag{B1}$$

For a combination of k series of measurements, if the variances are expected to be equal, a pooled standard deviation can be estimated from the sample standard deviations in order to give a better estimate of the variance. This can be calculated using equation B2 [B2].

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \ldots + (n_k-1)s_k^2}{n_1 + n_2 + \ldots + n_k - k}} \tag{B2}$$

Estimations of unknown population values can be made by using confidence intervals. A confidence interval is an interval, calculated from measured data, which is likely to include the true value. A confidence level is chosen, being a measure of the fraction of random samples for which the interval is correct. A common confidence level is 95%. Confidence intervals can be made for different parameters and can be written as:

$$sample\ estimate \pm multiplier * standard\ error \tag{B3}$$

The multiplier is a number that can be estimated from the confidence level by using a normal distribution or a t-distribution. A t-distribution is more spread than a normal distribution giving more probability in the extremes and is better to use when calculating a confidence interval based on a sample standard deviation.

A standard error is approximately the difference between the unknown population mean and the known sample mean. The standard error for a mean is calculated according to equation B4:

$$s.e.(\bar{x}) = \frac{s}{\sqrt{n}} s \tag{B4}$$

A confidence interval can also be made for a difference between two means. In this case the standard error becomes:

$$s.e.(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{B5}$$

If the variances are expected to be equal, the pooled standard error is used, represented in equation B6:

$$s.e.(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{(B6)}$$

When doing a confidence interval for the difference between two means there is an interest in knowing if the confidence interval estimate for μ_1-μ_2 includes 0. If it does, it is probable that the means are similar, if it does not they are probably different, depending on the confidence level used [B1].

References

[B1] Utts J. M., Heckard R. F. Mind on Statistics. 4th ed. Boston: Brooks/Cole Cengage Learning, 2011.

[B2] McNaught A. D., Wilkinson A. (ed). 2nd ed. IUPAC Compendium of Chemical Terminology. Oxford: Blackwell Science, 1997.

Appendix C: K_D determination protocol

Protocol describing the rounds performed with NMR for dissociation constant determination. The columns in table C1 describe round number, desired ligand concentration, the volume needed of 100 mM solution to reach the desired concentration, the volume to add each round of that solution, the volume needed of 10 mM solution to reach the desired concentration, the volume to add each round of that solution, the total volume received in the sample, the actual protein concentration and the actual ligand concentration.

Sample volume: 480 µl
Start conc.: 10 µM
Protein conc.: 2 µM

Table C1. Protocol describing amount of added ligand each round of NMR experiments.

Round	Ligand conc. (µM)	100 mM Volume (µl)	Incr	10 mM Volume (µl)	Incr2	Total volume (µl)	Protein conc. (µM)	Ligand conc. (µM)
1	10.00	0.05	0.05	0.48	**0.48**	480.48	2.00	9.99
2	24.00	0.12	0.07	1.15	**0.67**	481.15	2.00	23.94
3	57.60	0.28	0.16	2.76	**1.61**	482.76	1.99	57.27
4	138.24	0.66	0.39	6.64	**3.87**	486.64	1.97	136.36
5	331.78	1.59	**0.93**	15.93	9.29	487.56	1.97	326.63
6	796.26	3.82	**2.23**	38.22	22.30	489.79	1.96	780.34
7	1911.03	9.17	**5.35**	91.73	53.51	495.14	1.94	1852.58

For two ligands, 2271 and 6585, lower concentrations were needed and the protocol in table C2 was used.

Sample volume: 480 µl
Start conc.: 1 µM
Protein conc.: 2 µM

Table C2. Protocol describing amount of added ligand each round of NMR experiments for ligands 2271 and 6585.

Round	Ligand conc. (µM)	10 mM Volume (µl)	Incr	1 mM Volume (µl)	Incr2	Total volume (µl)	Protein conc. (µM)	Ligand conc. (µM)
1	1.00	0.05	0.05	0.48	**0.48**	480.48	2.00	1.00
2	2.40	0.12	0.07	1.15	**0.67**	481.15	2.00	2.39
3	5.76	0.28	0.16	2.76	**1.61**	482.76	1.99	5.73
4	13.82	0.66	0.39	6.64	**3.87**	486.64	1.97	13.64
5	33.18	1.59	**0.93**	15.93	9.29	487.56	1.97	32.66
6	79.63	3.82	**2.23**	38.22	22.30	489.79	1.96	78.03
7	191.10	9.17	**5.35**	91.73	53.51	495.14	1.94	185.26

Appendix D: Crystallographic and refinement details

In table D1 can be found crystallographic details of the different data sets in project B, including resolution, Mn(I/σ), Rmerge, completeness and R/Rfree.

Table D1. Crystallographic and refinement details for the data sets collected in project B. Numbers in brackets represent the values in the outer shell.

Protein/solvent	Resolution (Å)	MnI/σ	Rsym (%)	Completeness (%)	R/Rfree (%) partially refined
B1/H_2O	1.8 (1.85-1.80)	11.4 (3.7)	8.9 (38.9)	97.5 (94.0)	19.2/23.2
B1/DMSO	1.8 (1.85-1.80)	32.9 (15.9)	3.0 (8.0)	97.7 (92.6)	25.7/30.0 (deposited)
B1/Cyclodextrin	1.8 (1.85-1.80)	14.7 (4.6)	7.4 (33.8)	97.7 (98.3)	18.1/22.1
B1/DMA:PEG:H_2O					19.8/24.2
B1/nanosuspension, amorphous 1	1.8 (1.85-1.80)	19.9 (8.2)	5.2 (15.1)	93.8 (88.2)	19.2/22.4
B1, nanosuspension, crystalline 1	1.8 (1.85-1.80)	10.6 (3.5)	9.1 (40.4)	99.1 (94.9)	20.2/23.5
B1/nanosuspension, amorphous 2	1.8 (1.85-1.80)	17.4 (6.1)	5.6 (20.0)	99.7 (99.1)	19.3/23.5
B1, nanosuspension, crystalline 2	1.7 (1.74-1.70)	11.3 (2.8)	7.9 (34.6)	94.6 (73.8)	19.6/23.5
B2/DMSO	2.6 (2.67-2.60)	9.0 (2.0)	14.0 (67.7)	93.1 (94.5)	25.7/30.0

In table D2 can be found crystallographic details of the different data sets in project A, including resolution, Mn(I/σ), Rmerge, completeness and R/Rfree.

Table D2. Crystallographic and refinement details for the data sets collected in project A. Numbers in brackets represent values in the outer shell. Data were collected on a FR-E+ rotating anode generator, Saturn A200 CCD detector, Actor robot, λ=1.54 Å, with exceptions for [a] FR-E+ generator, R-axis HTC image plate detector, λ=1.54 Å [b] FR-E generator, R-axis HTC detector, Actor robot, λ=1.54 Å and [c] ESRF, ID23-1, λ=1.003 Å, [d] ESRF, ID23-1, λ=0.977 Å. Underlined R/Rfree values regards data sets for which the structure was not solved and which are excluded from the results.

Ligand/solvent	T (°C)	Resolution (Å)	MnI/σ	Rsym (%)	Completeness (%)	R/Rfree (%) partially refined
2271/H_2O	20	2.4 (2.56-2.42)	13.0 (2.9)	11.5 (58.2)	99.6(100)	23.6/29.0
2271/H2O[a]	30	2.3 (2.42-2.3)	7.8 (2.1)	9.2 (47.9)	91.7 (96.6)	27.8/31.4
2271/DMSO	20	2.3 (2.41-2.29)	15.4 (3.9)	8.9 (50.9)	99.8(100)	21.2/27.4
2271/Cyclodextrin	20	2.5 (2.56-2.50)	9.2 (2.0)	12.4 (67.3)	99.9 (91)	24.1/30.9
2271/DMA:PEG:H_2O	20	2.5 (2.67-2.54)	14.4 (3.5)	11.5 (53.5)	99.7(100)	23.3/30.0
6585/H_2O	30	2.3 (2.36-2.30)	11.5 (3.3)	7.3 (35.5)	93.1 (100)	23.4/28.9
6585/DMSO	20	2.4 (2.42-2.36)	8.7 (2.0)	8.3 (48.9)	98.3 (97.2)	24.6/27.7
6585/Cyclodextrin	30	2.3 (2.43-2.31)	14.5 (4.1)	11.0 (50.0)	99.9 (100)	20.4/25.3
6585/DMA:PEG:H_2O	20	2.4 (2.46-2.40)	4.6 (1.8)	20.4 (77.5)	92.1 (100)	28.2/38.3
6585/NaOH[c]	20	1.9 (2.05-1.94)	17.4 (2.8)	4.3 (40.1)	98.4 (98.3)	20.9/24.9
9692/H_2O	30	2.6 (2.70-2.56)	12.1 (3.0)	12.8 (54.4)	99.7 (99.9)	23.7/30.5
9692/DMSO	20	2.3 (2.41-2.28)	17.9 (5.4)	9.2 (39.0)	98.7 (99.6)	21.4/27.2
9692/DMSO	30	2.3 (2.38-2.26)	26.0 (9.7)	4.8 (17.6)	95.9 (91.1)	20.5/25.2
9692/DMSO, pH 8.4	20	1.8 (1.86-1.76)	20.6 (3.3)	5.4 (42.0)	96.9 (83.6)	24.3/27.3
9692/DMSO, pH 8.4	20	2.1 (2.19-2.08)	22.0 (5.7)	6.6 (29.7)	95.7 (99.7)	19.7/24.2
9692/Cyclodextrin[c]	30	2.0 (2.07-1.96)	15.0 (2.6)	5.1 (48.3)	99.7 (100)	19.7/24.1
9692/DMA	20	2.7 (2.77-2.70)	9.2 (2.1)	15.5 (72)	93.4 (100)	22.2/27.5
9692/DMA:PEG:H_2O[b]	20	2.7 (2.77-2.70)	9.2 (2.1)	15.5 (72)	93.4 (100)	21.4/27.2
9692/DMA:PEG:H_2O[c]	30	2.0 (2.07-1.97)	16.9 (3.2)	6.4 (51.6)	99.3 (99.1)	21.3/26.1
9692/HCl	30	2.7 (2.84-2.69)	15.6 (4.4)	14.9 (59.5)	94.2 (99.9)	21.0/29.4
1997/H_2O	30	2.3 (2.42-2.29)	13.6 (4.9)	10.4 (36.9)	99.6 (100)	20.9/26.3
1997/DMSO	30	2.4 (2.51-2.39)	22.8 (9.4)	6.7 (19.0)	99.7 (100)	20.7/26.1
1997/Cyclodextrin	20	2.2 (2.36-2.24)	17.9 (5.5)	8.3 (37.9)	99.6 (98.9)	23.1/27.0
1997/DMA:PEG:H_2O	20	2.3 (2.36-2.30)	9.8 (2.4)	8.5 (43.8)	92.6 (97.7)	29.4/33.5
1997/HCl	30	2.4 (2.41-2.35)	6.3 (1.9)	12.0 (53.3)	96.9 (95.5)	21.9/28.3
2349/H_2O	20	2.3 (2.41-2.28)	17.1 (3.8)	9.6 (49.7)	97.5 (96.0)	22.4/27.5
2349/DMSO	30	2.2 (2.33-2.21)	30.9 (9.2)	4.8 (21.5)	99.3 (99.3)	19.9/25.4
2349/Cyclodextrin	30	2.2 (2.32-2.20)	24.6 (7.0)	5.0 (25.5)	99.4 (98.2)	22.6/27.1
2349/DMA:PEG:H_2O	30	2.3 (2.40-2.28)	26.5 (8.5)	5.1 (24.0)	96.6 (94.5)	21.2/26.2
2349/HCl	20	2.3 (2.41-2.29)	22.4 (6.0)	7.9 (33.7)	97.8 (96.7)	20.9/26.7
9872/H_2O	20	2.3 (2.41-2.28)	26.1 (5.5)	5.4 (29.7)	99.9 (100)	20.9/25.7
9872/DMSO	20	2.3 (2.39-2.27)	23.8 (6.5)	6.2 (26.3)	97.5 (98.7)	18.9/24.1
9872/Cyclodextrin	30	2.3 (2.41-2.28)	24.7 (7.2)	5.2 (25.5)	99.4 (99.8)	22.1/27.7
9872/DMA:PEG:H_2O	30	2.3 (2.40-2.27)	30.9 (8.8)	4.1 (25.3)	99.0 (98.7)	20.5/25.0
9872/HCl	30	2.3 (2.38-2.26)	37.2 (11.3)	3.3 (14.6)	96.1 (94.1)	19.0/24.3
8246/H_2O	30	2.5 (2.61-2.48)	16.8 (3.7)	10.7 (53.9)	97.3 (95.7)	22.4/28.8
8246/DMSO	30	2.3 (2.39-2.27)	23.8 (5.7)	5.7 (32.1)	97.8 (99.8)	19.3/25.1
8246/Cyclodextrin	20	2.3 (2.46-2.33)	16.1 (3.4)	9.9 (51.8)	99.1 (99.9)	22.3/27.3
8246/DMA:PEG:H_2O	20	2.3 (2.40-2.28)	19.7 (4.9)	6.7 (34.5)	99.6 (100)	20.9/25.8
8246/HCl	30	2.3 (2.40-2.28)	20.3 (5.1)	6.5 (31.5)	98.3 (99.8)	20.7/26.6
6585/DMSO cocrystallization[d]	20	2.3 (2.41-2.29)	16.2 (3.6)	6.0 (43.2)	98.6 (99.5)	21.3/25.1

In table D3 is found further details regarding B1/DMSO as an example.

Table D3. Crystallographic and refinement details for B1/DMSO.

Space group	P212121
Cell parameters	a=27.58, b=83.96, c=103.17, α=β=γ=90.00
Number of molecules/asymmetric unit	2
Number of reflections	
Total	98450
Unique	22570
Resolution (Å)	24.66-1.80 (1.85-1.80)
Multiplicity	4.4 (4.3)
MnI/σ	32.9 (15.9)
Completeness (%)	97.7 (92.6)
Rmerge (%)	3.0 (8.0)
Mosaicity	0.64
R factor (%)	18.2
Free R factor (%)	21.2
Number of water molecules	196
r.m.s.d. (root mean square deviation) from ideal values	
Bond lengths (Å)	0.0071
Bond angles (°)	1.2149
Average B factors	
Main chain atoms	10.92
All protein atoms	13.10
Ligand	15.96
Water molecules	20.95
Metal ions	10.67

Appendix E: Repeated experiments

In table E1 can be seen measured solubility values for each sample for the experiments that were repeated. A sample mean and standard deviation were calculated in each case and used for further calculations of confidence intervals.

Table E1. List of repeated concentration determination experiments. Calculated concentrations from the different rounds and calculated mean values and standard deviations.

Combination	T (°C)	Exp 1	Exp 2	Exp 3	Mean	Standard deviation
2271/H_2O	20	63	50		57	9
	30	35	50		42	10
2271/DMSO	20	61	51		56	7
	30	42	49		45	5
9692/DMSO	20	34	37		36	2
	30	34	38		36	3
9692/HCl	20	49	50		49	1
	30	47	51		49	2
1997/H_2O	20	26	19		22	5
	30	25	20		23	4
1997/DMSO	20	46	46	53	48	4
	30	38		52	45	10
1997/Cyclodextrin	20	23	28	18	23	5
	30	40	27	25	30	8
1997/DMA:PEG:H_2O	20	37		29	33	5
	30	38	21	30	30	8
8246/Cyclodextrin	20	64	54		59	8
	30	52	46		49	4