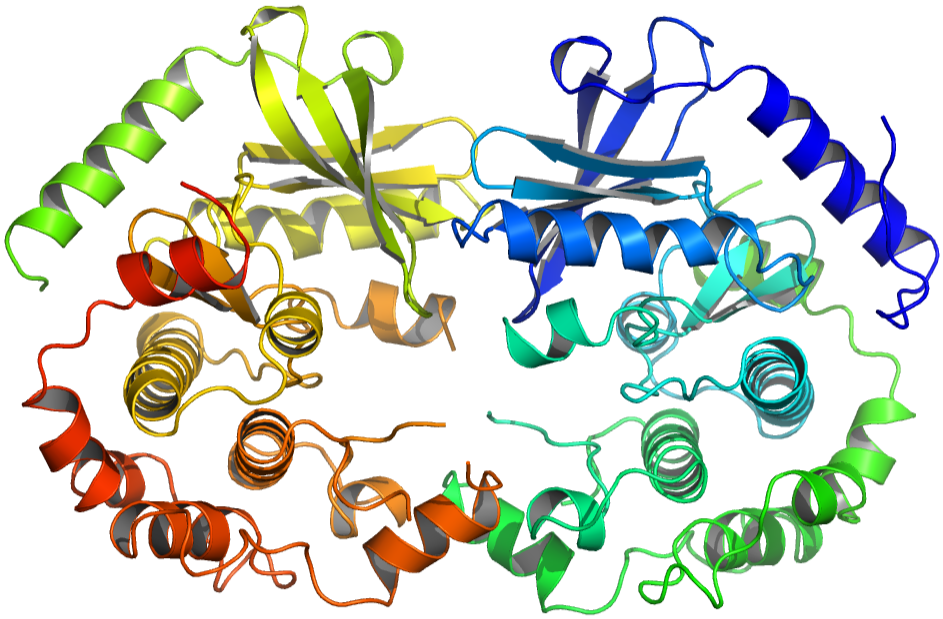
CHALMERS
UNIVERSITY OF TECHNOLOGY



X-ray fragment screening of MEK1 to identify allosteric fragment binders

CARL MÖLLER

Department of Biology and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2018

Master's thesis 2018

X-ray fragment screening of MEK1 to identify allosteric fragment binders

CARL MÖLLER

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Biology and Biological Engineering
Chemical Biology
Chalmers University of Technology
Gothenburg, Sweden 2018

X-ray fragment screening of MEK1 to identify allosteric fragment binders
Carl Möller

Supervisor: Helena Käck, Astra Zeneca
Linda Öster, Astra Zeneca
Examiner: Pernilla Wittung Stafshede, Chemical Biology

Master's Thesis 2018
Department of Biology and Biological Engineering
Division of Chemical Biology
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Representation of the MEK1 dimer configuration as it is found in the crystals used during this project.

Typeset in LaTeX
Gothenburg, Sweden 2018

X-ray fragment screening of MEK1 to identify allosteric fragment binders
Carl Möller
Department of Biology and Biological Engineering
Chalmers University of Technology

Abstract

The drug development process includes several phases and often span over many years. The initial phases are directed towards identifying and validating a potential target. Once a target has been validated, the hit identification and lead discovery process is initiated. This project have been performed on behalf of AstraZeneca and aimed to use x-ray crystallography to screen a compound library to identify fragments that binds to the allosteric pocket of MEK1. The results from this screen have mapped how a library of selected small molecules bound to the allosteric site. In total the campaign yielded 24 bound fragments with a good diversity that probes the entire pocket in a efficient way.

Contributors

This work was carried out with the support of the Diamond Light Source (proposal MX21007).

Acknowledgements

I would like to thank Linda Öster and Helena Käck for their support and patient explanations. I would like to thank all the people in the Structure Biophysics group in Discovery Sciences for much needed help. I would like to thank AstraZeneca for giving students the opportunity to work in a science and knowledge driven environment. Last but not least I would like to thank Linnea Johansson for the unwavering love and harsh comments, Johanna Ortgren for the long coffee breaks and all of my friends and family who supported me through out the project.

Carl Möller, Gothenburg, May 2018

Contents

List of Figures

List of Tables

1
Introduction

The drug development process includes several phases and often span over many years. The initial phases are directed towards identifying and validating a potential target. When a target has been identified its drugability is confirmed, e.g. if it is possible to inhibit or otherwise alter the target and if this will have a therapeutic effect. Once a target has been validated, the hit identification and lead discovery process is initiated. During this process libraries of compounds are systematically tested for members with the potential of giving the desired effect. This systematic testing can be done via several approaches. Either without any prior knowledge of the chemotype that is most likely to interact with the target. Or by the use of literature findings, statistical and structural data to create a subset of molecules with a appropriate chemotype[1].
This project have been performed on behalf of AstraZeneca with the aim to perform a primary x-ray crystallographic fragment screen on a validated drug target. The target is the dual-specific kinase MEK1 in the MAPK pathway and the results from this screen will map how a library of selected fragments binds to the allosteric site of MEK1. X-ray crystallography have been used to create a 3D atomic structure of the protein and a bound ligand. In a drug discovery setting the bound fragments could then have served as starting points for design of larger more potent compounds. Before the start of this project work had been initiated. This meant that there was already data present in some cases and that a working system for crystallisation and soaking of MEK1 had been established. The results from prior work will be presented along with the results from this project.

1.1 The Mitogen-Activated Protein Kinase pathway

The Mitogen-Activated Protein Kinase(MAPK) pathway is a signal transduction pathway involved in controlling intracellular events due to extracellular stimuli, such as growth factors or cytokines[2]. The mammalian MAPK pathway is conserved and organized in modules which contain kinases that are activated by phosphorylation cascades [2, 3, 4, 5]. The pathway is organized into four groups. The extracellular signal-regulated kinase(ERK), the c-Jun N-terminal kinase(JNK), p38 kinases and ERK5[3, 4, 6]. All four groups are activated by growth factors that stimulate cell surface receptors but JNK, p38 and ERK5 are additionally stimulated by environmental stress such as osmotic chock and ionizing radiation[6]. The four cascades, its members and their interactions are compiled in Figure 1.1.

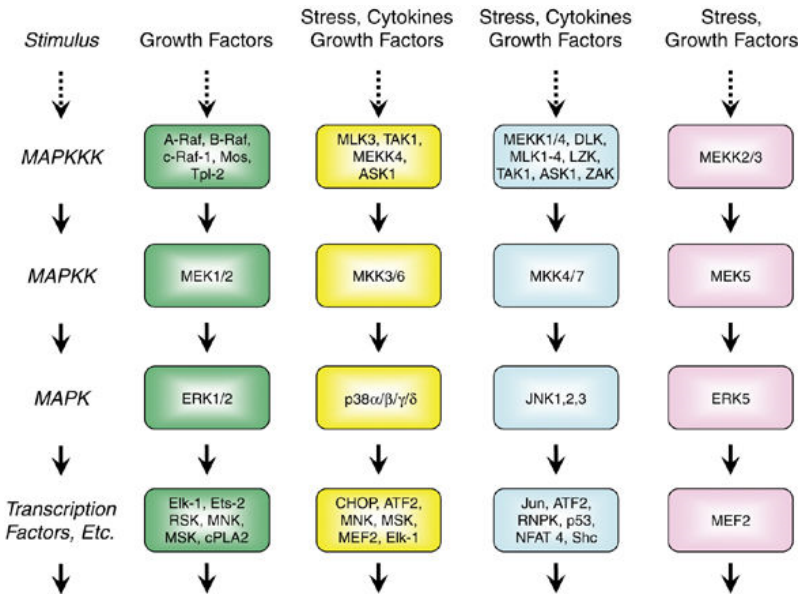


Figure 1.1: The extracellular signal-regulated kinase(ERK), the c-Jun N-terminal kinase(JNK), p38 kinases and ERK5 mammalian MAPK cascades. The figure is adapted from [6].

The ERK cascade have been shown to be responsible for the coordination and regulation of cell growth and differentiation due to extracellular stimuli as well as being associated with inflammatory response[4, 5]. Reports have also shown constitutive expression of phosphorylated ERK1/2 in tumours, indicating its role in tumour growth[7, 5].
Due to these observations it has been hypothesised that this cascade could be interrupted for therapeutic purposes. There are available drugs on the market that targets MEK1 such as cobimetinib[8] and trametinib[9].

1.2 Inhibition of kinases and its classification

Protein kinases are recognised to catalyse the reaction:

$$MgATP^{1-} + protein - O : H \Rightarrow protein - O : PO_3^{2-} + MgADP + H^+ [10]$$

The reaction phosphorylates a protein by replacing the H^+ of the OH-group with the gamma phosphate of ATP[11]. Since only a few aminoacids has a OH-group available protein kinases are divided in serine/threonine kinases or tyrosine kinases[10].
Kinases are involved in many key regulatory roles in cell biology and are a part of processes such as proliferation and apoptosis. Overexpression, mutation and dysregulation of kinases are directly connected to many diseases, many which are cancers[10, 11, 12]. This has made kinases a common drug targets and as of 2017 there were 35 FDA approved protein kinase inhibitors (PKIs)[12]. A majority of these are steady-state ATP competitive inhibitors interacting with the ATP pocket. Only three of the FDA approved drugs were allosteric PKIs targeting pockets outside the ATP-binding site [12, 10].

According to a review by Roskoski, kinase inhibitors can be classified into seven classes. Type I, I1/2, and II all refer to inhibitors that binds to the ATP-site but with subtle differences in the conformation. Type III and IV include allosteric inhibitors that are bound next to the ATP-site or bound away from the ATP-site respectively. Type V and VI describes bivalent and covalent inhibitors. Further subdivisions are possible and what criteria one employs in this subdivision differs but the general definition of the classes above are widely accepted.

1.2.1 MEK1

MEK1 is a dual-specificity threonine/tyrosine MAP/ERK kinase in the ERK cascade and phosphorylates the ERK protein[7]. Upstream it becomes phosphorlyated by RAF and has its phosphorylations sites at Ser218 and Ser222[8, 2]. In Figure 1.2 is MEK1 seen in a sideview with the N- and C-lobe marked out.

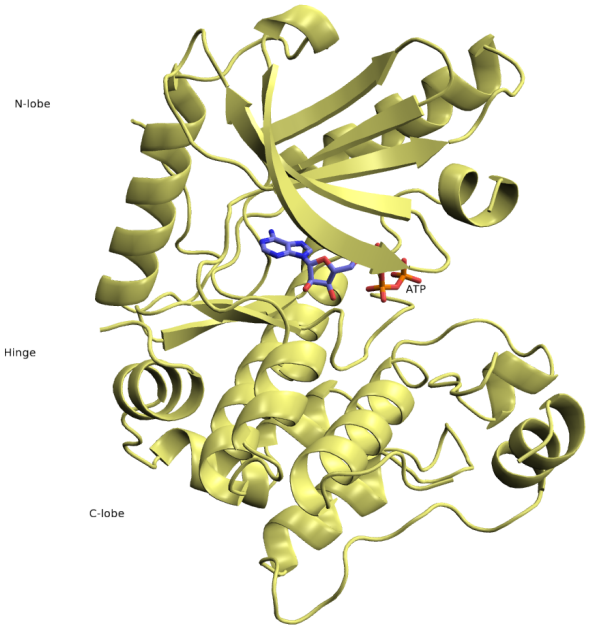


Figure 1.2: Overview of MEK1 with N- and C-lobe and hinge region marked out. The occupied ATP binding site is located in the hinge region with the allosteric pocket just to the right of it.

There are reports of MEK1 proteins that are adapted for crystallisation to form homodimers in both solution and the crystal[4]. The dimerization of an inactivated state of MEK1 has been dismissed to have any biological relevance[2].

1.2.2 Allosteric modulation of MEK1

Allosteric modulation is the regulation of an enzyme due to the binding of a modulator to another site then the active site. In most cases the binding of the modulator induces a change of the enzyme's conformation which in turn alters the binding affinity of the substrate[13]. In the case of kinases the binding of a allosteric modulator can entail changes that disables the kinase ability to phosphorylate or change the active site in such a way that it becomes unavailable. These changes can have either

a positive or negative impact on the activity which makes the modulator either a allosteric inhibitor or activator.

Many small molecule kinase inhibitors are ATP competitive, meaning that they inhibits the kinase by occupying the ATP-binding site. This can lead to problems with specificity, mutation resistance and the need for nanomolar affinities to compete with the intracellular levels of ATP[11]. Issues with ATP-competitive inihibitors that bind to other kinases then the one that is targeted is often avoided with allosteric inhibitors due to a relatively low sequence homology of allosteric sites which entails high specificity. In MEK1 the allosteric site is a pocket adjacent to the ATP binding site. In Figure 1.3 the pocket is shown with a surface. The entrance is located just to the right of the gamma phosphate in the bound ATP. The labelled residues are known interaction points from previously described allosteric inhibitors.

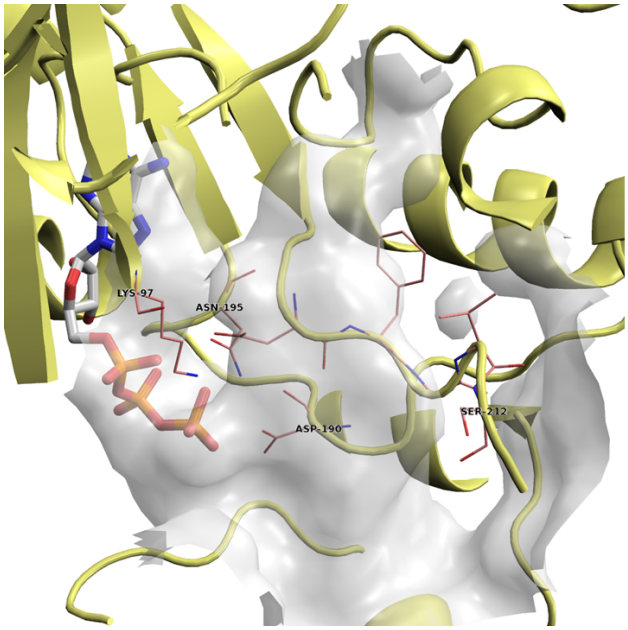
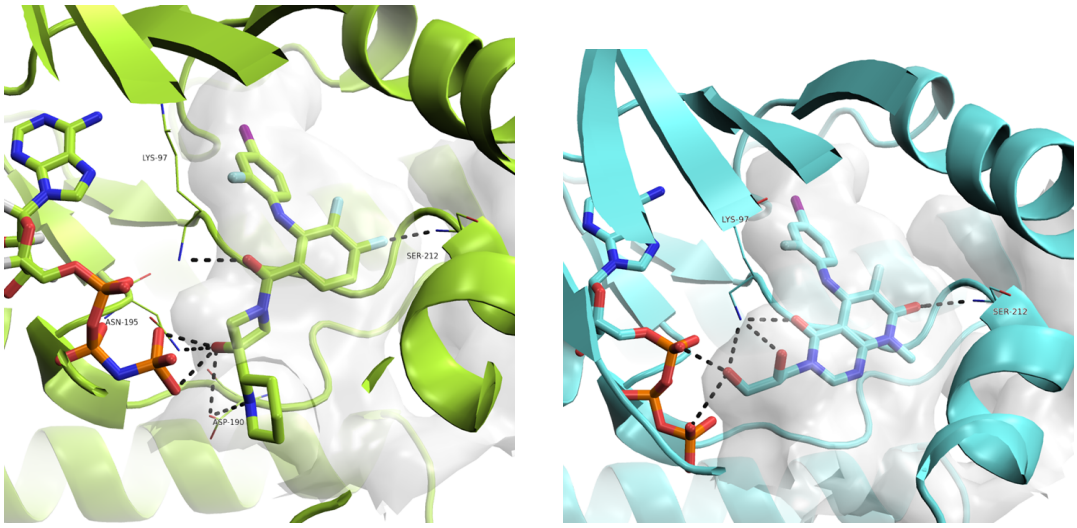


Figure 1.3: The allosteric site of MEK1 with previously described interaction points marked out.

1.2.2.1 Existing allosteric inhibitors of MEK1

Cobimetinib and Trametinib are FDA approved type III allosteric inhibitors. A crystal structure describing the binding of Cobimetinib to MEK1 has been published[8], however, no crystal structure is available in the literature for Trametinib. Another described MEK1 inhibitor is TAK-733 for which the complex crystal structure has been published[9]. Figure 1.4 shows the binding of Cobimetinib and TAK-733 to MEK1 respectively. The compounds both contains an aromatic ring binding in the deep part of the pocket, a central aromatic ring offering the potential to interact with Ser212 via substituents and an interaction with Lys97. They also expand towards the pocket entrance and interact directly with ATP and with two residues below the pocket[8, 14]. In Figure 1.4 are the crystal structures of cobimetinib and TAK-733 presented.



(a) Cobimetinib (PDB:4LMN,2.8Å) **(b)** TAK-733 (PDB:3PP1,2.7Å)

Figure 1.4: X-ray crystal structures of cobimetinib and TAK-733 binding to the allosteric site of MEK1.

1.2.3 Structure based drug design(SBDD)

The first structural information of proteins were retrieved in the 1950s and 1960s and it was at this time the idea of exploiting the targets structure to guide the development of drugs was first conceived[15]. The general idea is that if the structure of a target, with a ligand, compound or potential drug molecule bound, is known and understood, it is possible to design the chemistry in order to improve the binding properties of a ligand or inhibitor by optimising the fit of the ligand to the given pocket and potential specific interactions such as H-bonds or salt bridges.

1.2.3.1 Fragment based drug design(FBDD)

During the 1990s SBDD evolved and fragment-based drug design (FBDD) was developed. This approach probes the chemical space of a target using libraries of small molecules, referred to as fragments. Fragments are typically <300Da and a library size can range from a few hundred to several thousand fragments[16, 17]The concept is based on the fact that a combination of small fragments can sample the binding pocket of a protein more effectively than larger more drug like compounds, which are restrained by size[16].

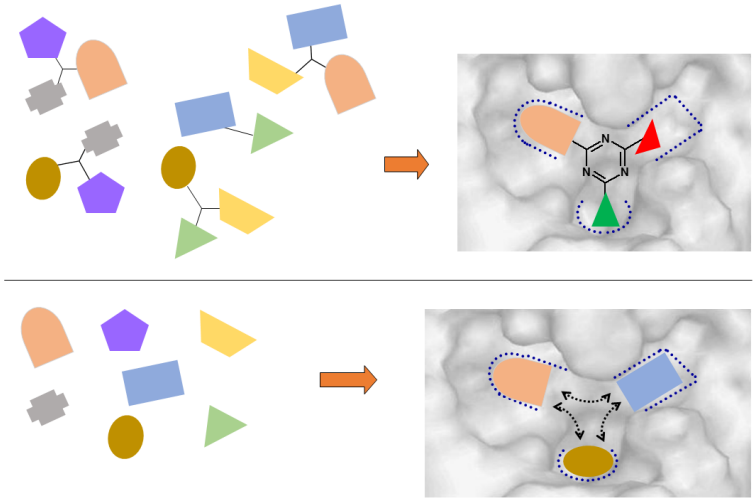


Figure 1.5: The top scenario shows how a larger compound from a HTS library has bound to the pocket. In the bottom scenario three separate hits from a FBDD library is shown and how they effectively have sampled the binding pocket. The fragment hits can be expanded or linked to create larger compounds with a good fit.

Figure 1.5 illustrates how FBDD differs from a high throughput screen(HTS) where larger compounds are used to generate starting points for lead generation. Early examples of FBDD made use of organic solvents that were soaked into crystals to demonstrate how small molecules could show potential ligand binding sites[15]. In a modern campaign carefully designed libraries of fragments are used in order to probe the chemical space in a systematic and rational way. Since small fragments typically bind weakly to the target protein due to few interactions they can be hard to measure in enzymatic assays. Therefore biophysical methods are normally used when trying to detect and characterise fragment binding, including Nuclear Magnetic Resonance(NMR), Surface Plasmon Resonance(SPR) and x-ray crystallography[15]. Once a set of binding fragments have been established and characterised they can be used to create larger compounds. As illustrated in the lower scenario of Figure 1.5 the fragments can be linked or expanded to form compounds with suitable functional groups.
X-ray crystallography holds many advantages as a primary screening method. It has a very wide affinity detection capability, ranging from sub-nanomolar to millimolar, that is only limited by ligand solubility. It is also close to unlimited when it comes to size of the protein and depend only on crystal stability and resolution limit. In addition a 3D structure describing the binding of the fragment to the target gives important information when expanding or linking fragments. The most common objection is the impracticality of a crystallisation system and the labour intense and time consuming process of setting up such a system[16].

1.2.4 X-ray crystallography and fragment screening

A fragment screening campaign based on detection with x-ray crystallography consists of several steps which needs to be properly designed and optimised. The crys-

tallisation of the protein is usually a trial and error process where different solutions are tested and optimised[18]. When the fragments has been introduced to the protein crystal coherent x-rays are directed towards the crystal and a diffraction pattern is obtained. From this pattern an electron density map can be calculated, which is a representation of the atoms that constitute the crystal. The electron density map is then used to place atoms so that they fit with the observed density and creates a 3D structure of the protein and potential bound fragments. In Figure 1.6 a generalised pipeline is presented with the steps necessary to be able to build the 3D atomic models from which the presence of a bound fragment is determined.

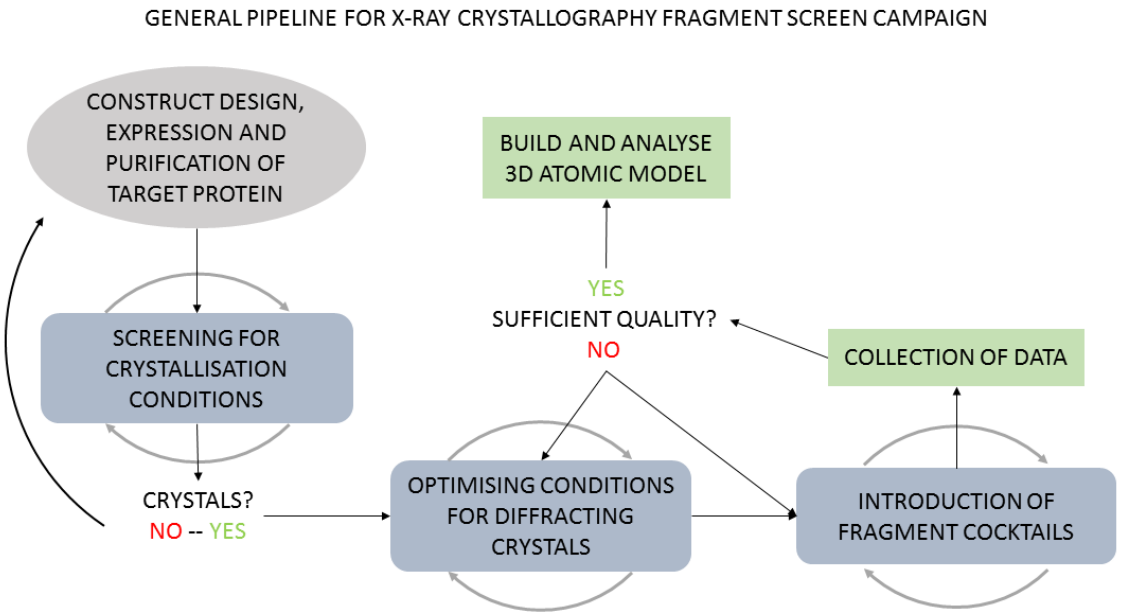


Figure 1.6: The flowchart describes a generalised process of a x-ray crystallography fragment screening campaign.

Aim of the study

The project aim is to use x-ray crystallography to screen the allosteric pocket of MEK1 to identify fragment binders that can be used as alternative chemical starting points.

2
Theory

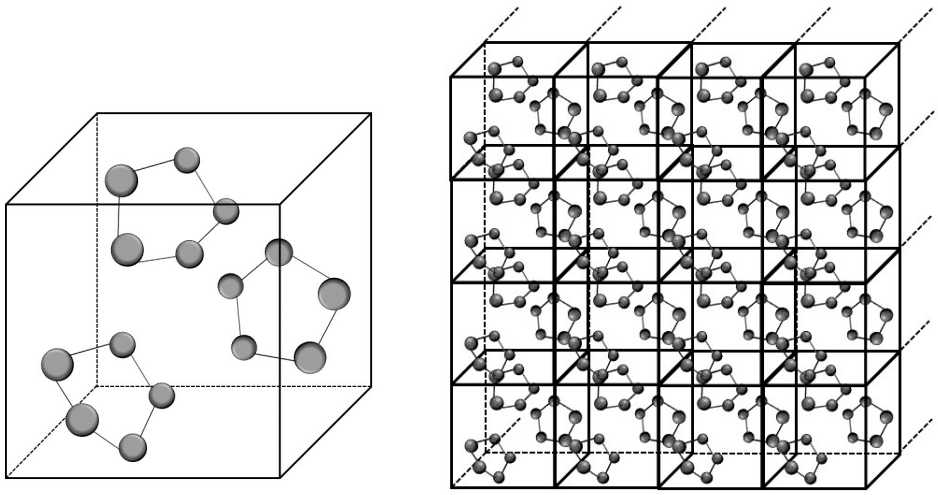
2.1 X-ray crystallography

X-ray crystallography aims to determine the atomic structure of a protein by looking at how a x-ray beam travels through a crystal and diffract. The diffracted waves are recorded by a detector. From the recorded pattern it can be calculated were and how the incoming wave have been scattered. From this the electron density can be derived to which a model of the protein is fitted.

2.1.1 Protein crystallisation

Good quality crystals are essential in order to get high resolution data. What drives the formation and growth of crystals are well understood, but how to induce it less so. A large part of the crystallisation procedure is still trial and error and rely on the experience of the crystalliser.
The event that initiates the growth of a crystal is called nucleation[19]. This event occurs due to either the presence of solid particles in the solution or spontaneously by the protein one wishes to crystallise[20]. In the first scenario the macro molecules simply adsorb to the particle, which can be dust or other impurities, forming a stable core from which the crystal can grow. Spontaneous nucleation occurs when the protein, in solution, form clusters. Under a certain level of supersaturation these clusters tend to dissociate but if the level of supersaturation is great enough this tendency is overcome by the deposition of protein on the cluster and nucleation occurs[20]. Crystal growth then occurs by the localisation of a protein to a suitable site on the crystal surface followed by its incorporation[19].
A crystalline form is defined by the orderly three dimensional repeated arrangement of molecules held together by noncovalent interactions[21]. Compared to inorganic and small molecule crystals protein crystals are very fragile. This is due to that protein crystals often rely on weaker forces like hydrogen bonds rather then electrostatic forces of fully charged ions. This together with relatively low number of interactions creates a fragile structure [21, 18]. Due to the size and shape of macro molecules the crystals that are formed are often porous which allows for a high solvent content. Approximately 50% of the crystal volume is solvent but can vary from 25% to 90% in some cases [18]. This gives the possibility to introduce compounds to proteins while in crystalline form and for the protein to remain active.
A schematic of a unit cell with arrangement of macro molecules is shown in Figure 2.1a with a view of the ordered repeating arrangement in Figure 2.1b. Each box represent what is called a unit cell which are identical arrangement of molecules[21].

The array presented in Figure 2.1b is in reality not present throughout the crystal.



(a) Schematic of a unit cell. **(b)** Schematic of the crystalline state

Figure 2.1: (a)A unit cell is the smallest symmetric unit. It can contain several macro molecules. (b) A crystal is a symmetric three dimensional repetition of the unit cell.

Instead the crystal is composed by small subarrays slightly offset to each other. This gives the crystal a property called *mosaicity* which results in the divergence of the diffracted x-ray beam[21].

2.1.2 Crystallisation

The main issue when trying to grow protein crystals is to find conditions under which crystallisation occurs. To induce this event an aqueous buffer called mother solution is created, consisting of well known quantities of precipitants. The purpose of this solution is to create tolerable conditions where the protein is stable. There is no usable theory that predicts what conditions that will yield crystals, instead an empirical approach needs to be taken. This often leads to large screening trials with several rounds of optimisation before quality crystals are obtained[22, 23, 18, 21].

2.1.2.1 Nucleation and Supersaturation

The nucleation is the critical event where the molecules proceed from being in a fully disordered state to an ordered. It is believed that this occurs in steps of partially ordered aggregates with short range order that eventually reaches a fully ordered state, called a critical nuclei. Once this is established the crystal can start to grow[18]. In order to initiate these nucleation events the protein needs to precipitate out of solution. This is done by creating a supersaturated state. Supersaturation is a non-equilibrium state where a part of the soluble protein is in excess of what the present conditions allow. To reestablish equilibrium the excess part of protein forms a solid state, i.e. crystals. To produce this supersaturated state one needs to slowly change the conditions of the solution such that the solubility limit changes. In Figure

2.2 a general solubility curve is presented showing how the supersaturated state occurs as the protein and adjustable parameter, i.e. salt concentration, increases.

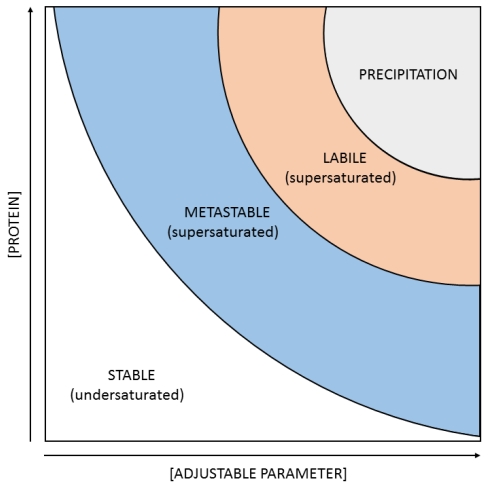


Figure 2.2: Schematic overview of how the supersaturated states changes with protein concentration and adjustment of parameters. Depending on the protein and the parameter the relation will look different. Image adapted from [18]

2.1.2.2 Precipitants, physio- and biochemical parameters

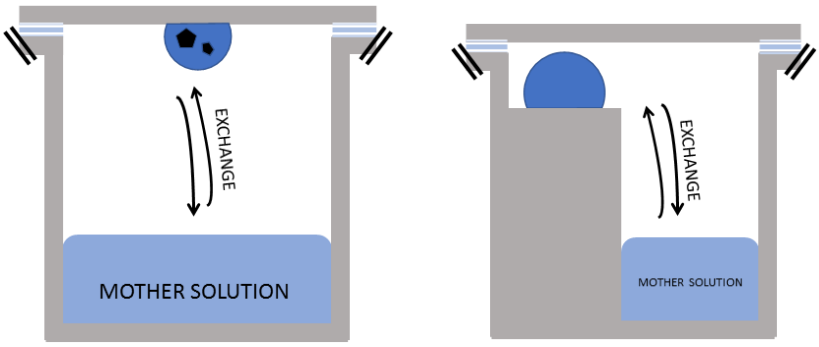
The number of factors that affect crystal growth are many. In Table 2.1 a selection of physical, chemical and biochemical factors is presented. For practical reasons most parameters are commonly fixed.

Physical	*Chemical*	*Biochemical*
Temperature	pH	Ligands
Pressure	Ionic strength	Purity
Rate of equilibirum	Protein concentration	Proteolysis
Time	Precipitant	Unstructured regions
Viscosity	Surfactants	Conformational states

Table 2.1: Factors that affect crystallisation. Table adapted from [18]

2.1.2.3 Vapour diffusion

Vapour diffusion is the most common technique to induce the supersaturated state. A drop consisting of equal parts mother solution and pure protein is placed in a sealed environment with a reservoir of mother solution. This will enable water, or volatile precipitants, to diffuse between the reservoir and the drop, thus altering the concentration of precipitants and protein in the drop. If the conditions are right a supersaturated state will occur and crystals grow. In Figure 2.3 are two different variants of vapour diffusion presented. The hanging drop technique allows for easy handling and to not demand special plates. A sitting drop approach makes larger drops possible and allows for the use of precipitants that lower the surface tension[23]. In Figure 2.3b the drop is placed on a plateau in order to simplify the use of robotics and submicrolitre dispensing.



(a) Hanging drop (b) Sitting drop

Figure 2.3: Schematic view of a setup with a **(a)**, hanging and **(b)**, sitting drop. In this case the sitting drop is placed on a plateau in order simplify for robotic dispensers. As water is transported via vapour diffusion the concentrations changes in the reservoir and the drop.

2.1.3 Diffraction of x-ray waves

Diffraction happens when a wave hits an obstacle and changes direction or are scattered in many directions. When a x-ray beam travels through a protein crystal the scattering occurs due to an interaction between the x-ray, as an electromagnetic wave, and the electrons. The electric and magnetic components of the wave exerts a force on the electron who in turn starts to oscillate. This oscillation emits radiation in the same frequency as the incident wave but will do so in all directions[22]. This kind of scattering is called elastic scattering due to the preservation of the kinetic energy. In the case of x-ray diffraction the electrons can be considered as free electrons and are thus not affected by the forces exerted by the nucleus. This implies that if the scattered beam can be tracked to the point were it was scattered and it is known how the intensity of the scattered wave changed, the electron density at that point can be derived[22].
The pattern that occurs when recording the diffracted x-ray beams from an arbitrary angle typically look like the pattern in Figure 2.4.

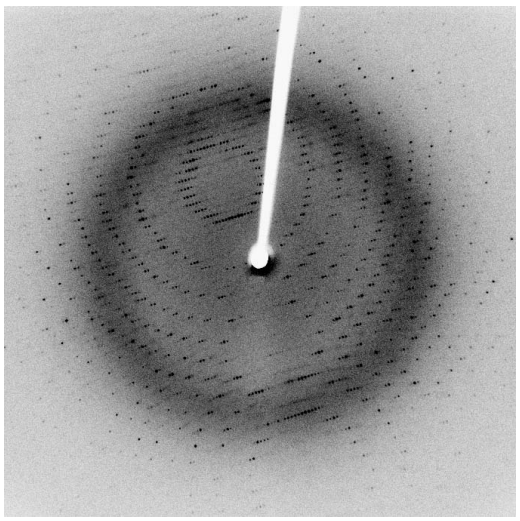


Figure 2.4: A typical diffraction pattern from arbitrary angle. The white are is the centre of the beam and the beam stop. Image adapted from [24]

The very distinct pattern is a representation of the reciprocal lattice and is explained by Bragg's law and the condition that needs to be met in order for constructive interference to occur e.g. the addition of waves.

2.1.3.1 The Unit Cell

The unit cell is defined by the vectors **a**, **b** and **c** with the angle γ between ab, angle β between ac and angle α between bc. The **a**, **b** and **c** defines the crystal lattice x, y and z directions respectively[21, 22, 25]. This allows for the jth atoms position in the unit cell to be described by the fractional coordinates (x_j, y_j, z_j). Which are in the range 0-1[25].

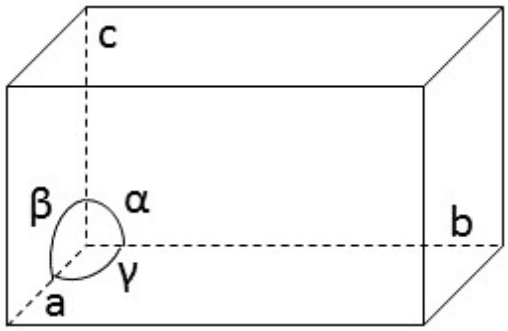


Figure 2.5: The unit cell is defined by the vectors **a**,**b** and **c** with the angle γ between ab, angle β between ac and angle α between bc[21, 25, 22].

As depicted in Figure 2.1b a crystal is a three dimensional stack of unit cells. These will together create a lattice with lattice points along the faces of the unit cells[21, 22]. These lattice points will define a set of planes with equal distance to each other, as depicted in Figure 2.6. Depending on the shape and size of the unit cell a number of these sets can be constructed. As described and explained in subsection 2.1.3 these planes are very important for our understanding of where and why diffraction is produced[21, 22].

2.1.3.1.1 Space group Depending on how many molecules that are present inside the unit cell degrees of symmetry can be present. This symmetry is referred to as a space group and describes rotational and translational symmetry within the unit cell. The largest group of molecules that has no internal symmetry is called asymmetric unit and can be fitted to equivalent positions due to symmetry operations. An example of a space group is $P2_12_12_1$ which describes a space group with three perpendicular two-fold rotational axes[21, 22]. The symmetry of the unit cell is also present in the reciprocal lattice, see section 2.1.3.3, which means that some reflections are identical. This in turn means that it is possible to record all reflections without having to turn the crystal 360°[21, 22].

2.1.3.2 Bragg's law

Lawrence Bragg and his father William Bragg got the Nobel prize in Physics in 1915 for the interpretation of x-ray diffraction and the formulation of the conditions that need to be met in order for constructive interference to occur[25]. This condition is known as Bragg's law, see Equation 2.1, and tells us that constructive interference

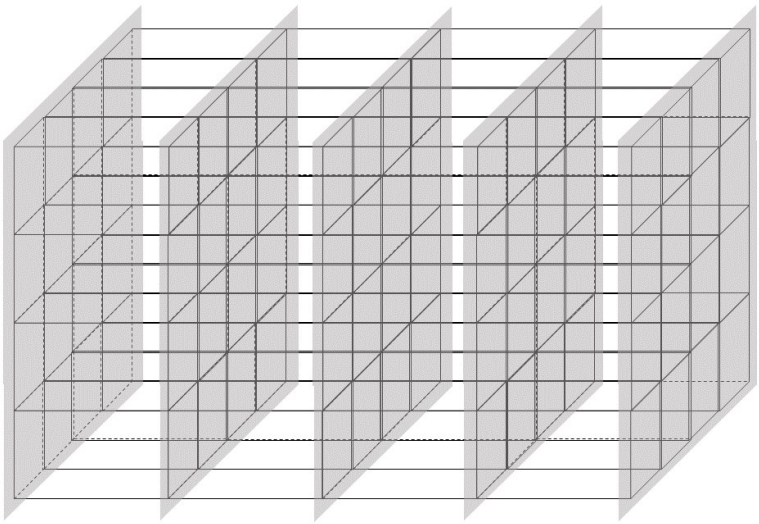


Figure 2.6: Through the lattice points defined along the vertices of the unit cell many set of planes can be defined. This is exemplified by the set of planes defined by the **a** and **c** vectors of the unit cell. All planes are equally distanced, $d_{hkl} = \mathbf{b}$, apart.

will only happen if the difference in path length(2BC) of two parallel waves are a whole number of wavelengths(nλ)[25, 22, 21].

$$2d_{hkl}sin\phi = n\lambda \tag{2.1}$$

Lawrence Bragg described the diffraction of the x-ray in the crystal as a reflection of a richly populated plane of atoms. In Figure 2.7 two such planes are depicted. The incident waves have the same angle(ϕ) and wavelength(λ). In order for the waves to be scattered in phase the distance between these planes need to meet the condition described in Equation 2.1.

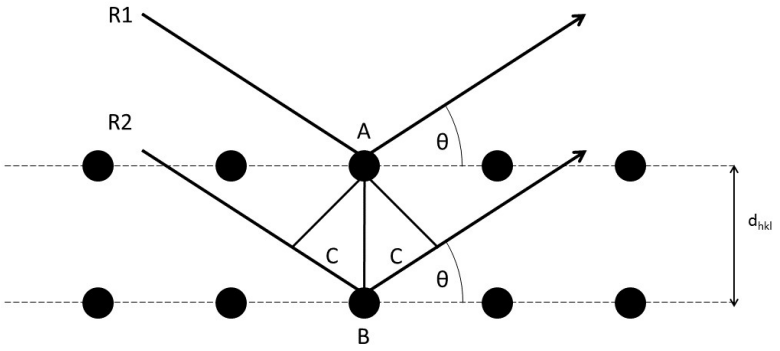


Figure 2.7: Each point is a lattice point and the two rows represent a plane in the lattice with the interplanar distance d_{hkl}. Image adapted from [21]

If this condition is not met the waves will be scattered out of phase and destructive interference will occur. This explains the distinct pattern of the diffracted waves. See Figure 2.4.

2.1.3.3 The reciprocal lattice and the sphere of reflection

The reciprocal lattice is the crystal lattice represented in reciprocal space[21]. The lattices relates in that each set of planes in the crystal lattice is represented by a point in the reciprocal lattice[26]. The reciprocal lattice is an imaginary construction that allows for a simple way to determine the direction of the scattering[22]. In Figure 2.4 it is clearly seen that the reflections follow a expanding circular pattern. This can be explained by Ewald's sphere, or the "sphere of reflection" as it is also refereed to[26]. The sphere has its origin in the crystal and has a radius of $1/\lambda$. Like in the crystal lattice a set of planes can be constructed through the lattices points of the reciprocal lattice. These planes will intersect with the sphere, and if a lattice point is located at this intersection diffraction will occur[22, 26, 21]. This can be rephrased as the event of Bragg's law being fulfilled so that constructive interference occurs and thus the addition of diffracted waves.

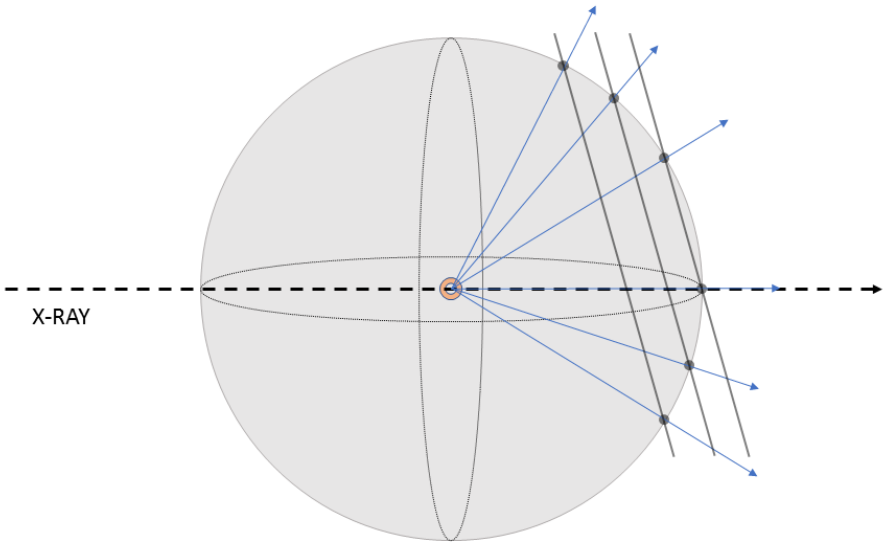


Figure 2.8: The sphere of reflection(Ewald's sphere) has its origin in the crystal and has a radius of $1/\lambda$. Like in the crystal lattice a set of planes can be constructed through the lattices points of the reciprocal lattice. These planes will intersect with the sphere, and if a lattice point is located at this intersection diffraction will occur. This can be rephrased as the event of Bragg's law being fulfilled so that constructive interference occurs and thus the addition of diffracted waves.

This means that the recorded diffraction pattern is a 2D representation of the reciprocal lattice. And since the reciprocal lattice is fixed in relation to the crystal lattice this will rotate as we rotate the crystal which allows us to record the reciprocal lattice from a series of angles and thus determining the reciprocal lattice 3D structure. Since we know that all points in the reciprocal lattice is located on, and where on, the sphere of reflection, and the incoming angle of the waves are known, it can be calculated in what direction the waves have been diffracted.

2.1.3.3.1 Reciprocal and real space coordinates The 3D reciprocal lattice is indexed according to a coordinate system using the indices hkl. hkl are integers since the reflections are indexed by counting from $hkl = 000$. This is placed at the central reflection of the reciprocal lattice, e.g. the point where the x-rays pass straight through the crystal. As mentioned in subsubsection 2.1.3.3 the position of the reflection is uniquely specified by the angle of divergence. In the crystal lattice(real space) an atoms position is described by xyz. Distances are given in Å in both spaces with the difference that it is inverted in reciprocal space, 1/Å. This means that if a outer reflection is placed 1/xÅ from $hkl = 000$ one can expect to get a model with a resolution of xÅ[21].

2.1.4 From recorded reflection to electron density

2.1.4.0.1 Structure factor Since each recorded reflection is a point in the reciprocal lattice it can be described by a Fourier series. A Fourier series that describes a diffracted x-ray is called a *structure-factor* and is denoted with F_{hkl}. The fact that it is referred to as a *series* suggests that it is a sum. A structure-factor is the sum of all the diffractive contributions given by the N atoms in the unit cell. The structure factor can be written as Equation 2.2[21].

$$F_{hkl} = \sum_{j=1}^{N} f_j \, e^{2\pi i(hx_j + ky_j + lz_j)} \tag{2.2}$$

f_j is the scattering factor for each individual atom which allows it to be treated as a single sphere of electrons. This factor is unique for every element depending on the number of electrons. This means in words that the contribution from each atom to the structure factor depends on the kind of element it is, f_j, which determines the amplitude of the contribution, and secondly its position in the unit cell (x_j, y_j, z_j), which determines the phase of its contribution[21].

2.1.4.0.2 Electron density Due to the ordered nature of crystals and the identical arrangement of atoms in every unit cell the electron density function, $\rho(x, y, z)$, that describes the surface of the electron cloud surrounding every atom is a periodic function. If $\rho(x, y, z)$ is plotted it will yield the 3D shape of the electron cloud and thus the 3D shape of the molecule surrounded by that cloud[21]. Since f_j in Equation 2.2 also can be defined as the average electron density in an infinitesimal volume element of the unit cell the structure factor can be alternatively written as Equation 2.3[21].

$$F_{hkl} = \int_V \rho(x, y, z) \, e^{2\pi i(hx + ky + lz)} \, \mathrm{d}V \tag{2.3}$$

This implies that the structure factor described in paragraph 2.1.4.0.1 can be used to calculate the electron density function since the Fourier transform is reversible. This property gives us Equation 2.4. Instead of an integral it is a triple sum since hkl are indices.

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l F_{hkl}\, e^{-2\pi i(hx+ky+lz)} \tag{2.4}$$

The expression given in Equation 2.4 allows for the electron density to be calculated if we can construct the Fourier series from the structure factors F_{hkl}. F_{hkl} describes the wave equation of the diffracted x-ray. This means that F_{hkl} must specify frequency, amplitude and phase. The amplitude is proportional to $(I_{hkl})^{1/2}$ and is given by the recorded reflections. The frequency is the same as the x-ray source and the phase is unknown. This is called the phase problem and it cannot be determined directly from the recorded reflections[21]. To obtain the phases one can use previously solved structures that are assumed to be similar to the structure one trying to obtain. This is called molecular replacement.

2.1.4.0.3 Molecular replacement Molecular replacement assumes that the phases of an unknown structure is similar or identical to the phases of a structure which is already solved. The solved structure can thus be used as *phasing model* to solve the phase problem every time a new ligand is introduced by assuming that the phase is the same as for a structure without a bound ligand. By taking the phases of the known structure, $\rho(x,y,z)$ can be calculated for the model without known phases. This means that, since the phase is related to the position of every atom in the unit cell, the known molecular structure needs to be positioned in the unit cell of the unkown structure in such a way that it will give phases most like the phases of the unknown structure. Hence *molecular replacement*. One way to find this orientation is to minimise the result of Equation 2.5, which simply compares the agreement between the two sets of structure factors. The observed structure factors of the unknown structure and the calculated structure factors obtained when fitting the known structure[21, 22].

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|} \tag{2.5}$$

This method is also applicable in cases where the structure of a homologous protein is known. Since proteins with a similar amino acid sequence are likely to have a similar fold, homologs can be a way to solve the phase problem for previously unsolved structures[21, 22].

2.1.4.0.4 Refinement Refinement is the process where the model parameters are improved in order to explain the experimental data as well as possible. The most common approach is the maximum likelihood principle which assumes that if the model is correct the likelihood to observe a structure factor equal to the one of the model should be high. If $P(|F_o|\,;|F_c|)$ is the probability of observing a structure factor F_o if we have a calculated structure factor F_c then the joint probability can be conveniently written as Equation 2.6. The refinement then become a matter of minimising L[27, 28].

$$L = -\sum_{hkl} log(P(|F_o|\,;|F_c|) \tag{2.6}$$

3

Methods

3.1 Fragment screening using X-ray crystallography

A fragment screen by x-ray crystallography was performed by soaking crystals with a AstraZeneca inhouse-designed fragment library. There were in total 96 cocktails containing 3-5 fragments each. Each cocktail will be referred to by its position in the 96 well plate where A-H refers to the row and 1-12 to the column. Crystals was grown according to a protocol designed previous to this project and several crystallisation conditions were tested. The structures was solved and analysed for fragment binding in the allosteric pocket of MEK1 by visual inspection of electron density areas which was not occupied by the protein model.

3.1.1 Protein production and purification

The protein used for all experiments was kindly provided by L. Barlind(AstraZeneca) and sourced from the same batch. Expression was done in insect cells with a construct(MEK1 37-383) containing a His-tag situated in a manner not interfering with the dimer interface. The protein had a calculated molecular weight of 36977.55 (g/-mol), a extinction coefficient of 24410 (M-1 cm-1 at 280 nm). The concentration was measured to 9.8mg/ml[29].

3.1.2 Crystallisation and introduction of fragments

Crystallisation was achieved with the hanging-drop technique in a 24 well VDXm Plate with sealant from Hampton Research. Each well was filled with $500\mu L$ of mother solution. Each drop was constituted by $1\mu L$ mother solution and $1\mu L$ of protein solution. The calculated protein concentration in each drop was 0.135 mM. The concentration of salt and and pH in the mother solution was varied along the rows and columns respectively, see Figure 3.1 for plate layout. All components of the mother solution and cryo-solutions was provided by Hampton Research.

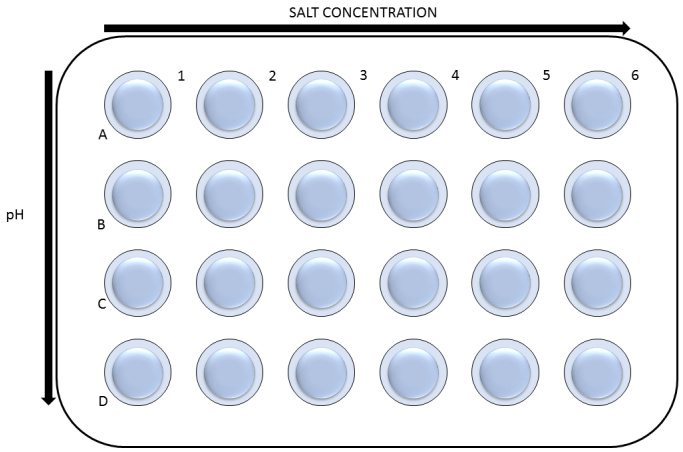


Figure 3.1: Illustration of the 24 well plate used for the hanging drop technique and how its coordinates are defined. In the different plate designs the salt concentration was varied along the rows and the pH along the columns. See Table 3.1-3.5 for details.

3.1.2.1 General protocol

Through the project the same general procedure has been used when preparing the protein and mother solution for crystallisation with the hanging drop technique. The nonhydrolyzable ATP-analog adenosine 5'-(β,γ-imido)triphosphate(AMP-PNP) was added to avoid fragment binding in the ATP binding site and provide the possibility for fragments to interact.

1. Defrost protein aliqout
2. Centrifuge at high speed for 10 min in order to sediment precipitate
3. Transfer to new tube, add AMP-PNP[2mM] and $MgCl_2$[2mM]
4. Incubate for minimum of 30 min at room temperature
5. Centrifuge at high speed for 20 min in order to sediment precipitate
6. Create a $1\mu L$ droplet on a glass slide and add $1\mu L$ mother solution

Given 9.8 mg/ml of protein in the stock solution, corresponding to 0.265M, each drop had a final calculated concentration of 0.135 mM.

3.1.2.1.1 Mother solution The mother solution used was constituted by varying concentrations of salt solutions and solutions with buffer solutions with varying pH.

- Tris - $(HOCH_2)_3CNH_2$
- Bis-tris - $C_8H_{19}NO_5$
- Ammonium Sulphate - $(NH_4)_2SO_4$
- Sodium Chloride - $NaCl$

To prepare each plate a liquid dispensing robot, Formulator, was used. The Formulator introduced each component directly to the well after which the plate was placed on a shaker in order to ensure a homogeneous solution. The droplet was dispensed manually on the silicone coated glass slide.

3.1.3 Evaluation of tested crystals and crystallisation conditions

Since well diffracting crystals was fundamental to be able to determine fragment presence the tested crystals were evaluated based on the quality of the data that they generated. Good quality data was defined as data that generated crystal structures with high enough quality to define if a fragment was bound or not. The crystals that yielded data sets annotated as *good* were mapped to the different conditions. By ranking the conditions based on how many good quality data sets the crystals grown in them had generated new parameter values could be identified. In a initial step the tested crystals were also evaluated based on their appearance and common features were noted in order to improve the selection process during harvesting. This initial evaluation was done visually from images taken on the loop during data collection. See Figure A.15 for a representative example.

During the project the crystallisation conditions were changed according to five different designs. The plate designs that were used, are described in Table 3.1-3.5.

Precipitant	Concentration	pH	Wells
Ammonium Sulphate	1.9-2.2M		A1-D6
Sodium Chloride	0.1 M		A1-D6
Tris	0.1M	7-7.4	B1-D6
Bis-Tris	0.1M	6	A1-A6

Table 3.1: Design 1: Initial conditions designed previously to the start of the project.

Precipitant	Concentration	pH	Wells
Ammonium Sulphate	1.9-2.1M		A1-D6
Sodium Chloride	0.1 M		A1-D6
Tris	0.1M	7-7.4	B1-D6
Bis-Tris	0.1M	6	A1-A6

Table 3.2: Design 2: The ammonium sulphate concentration ranged were narrowed since more crystals appeared in the wells with lower concetration.

Precipitant	Concentration	pH	Wells
Ammonium Sulphate	1.85-2.05M		A1-D6
Sodium Chloride	0.1 M		A1-D6
Tris	0.1M	7-7.3	B1-D6
Bis-Tris	0.1M	6	A1-A6

Table 3.3: Design 3: The ammonium sulphate concentration range were lowered to find a lower limit and decrease the speed of crystal growth. The tris pH range were lowered due to more crystals in the lower end of the range.

Precipitant	Concentration	pH	Wells
Ammonium Sulphate	1.9-2.2M		A1-D6
Sodium Chloride	0.1M M		A1-D6
Bis-Tris	0.1M	5.5-7.5	A1-D6

Table 3.4: Design 4: The crystal yield from Design 1 and 2 indicated that wells with only Bis-tris instead of the combination with Tris yielded more crystals. The pH range were widened and the ammonium sulphate concentration reset to initial conditions.

Precipitant	Concentration	pH	Wells
Ammonium Sulphate	2.01-2.03M		A1-D6
Sodium Chloride	0.1M M		A1-D6
Bis-Tris	0.1M	6.1-6.24	A1-D6

Table 3.5: Design 5: Both the ammonium sulphate concentration range and Bis-tris pH range were narrowed to optimise the conditions with the best yield.

3.1.3.1 Soaking and freezing

The different fragment cocktails was introduced to the crystals via soaking according to the following steps. Note that the concentrations given describes the concentration of each individual fragment.

1. The cocktails were dissolved in DMSO to a stock of 500 mM.
2. Further dilution to 75 mM of fragments with a solvent consisting of the same components as the mother solution.
3. Transfer of $1\mu L$ 75mM solution to a drop containing a crystal, giving a fragment concentration of 25mM
4. Incubation for 2h

.

The protein concentration in the drop after addition of fragments was 0.088mM. After the addition of $1\mu L$ cocktail solution to the $2\mu L$ hanging drop the concentration of each fragment was diluted to 25mM, giving close to a 1:280 relation between the protein and fragment concentration.

	Concentration
Ammonium Sulphate	3M
Sodium Chloride	0.1M
Tris pH 7	50mM
Bis-Tris pH 6	50mM

Table 3.6: Cryo-solution used when freezing crystals.

After incubation the crystal were extracted from the drop with a nylon loop under a microscope, subsequently dipped in a cryo-solution and submerged in liquid nitrogen. The cryo-solution(3M $AmSO_4$, 0.1M NaCl, 50mM Tris pH7, 50mM Bis-Tris pH6) was constituted by the same components as the mother solution but adjusted to avoid formation of ice crystals when freezing the crystals.
In the cases where no good data could be collected the soaking was repeated following the same protocol. If a no usable data could be collected after the second try it was

assumed that a fragment, or a combination of fragments, deteriorated the crystal. During the screen there existed cases where the cocktail had to be convoluted, e.g. the fragments had to be introduced one by one in order to determine which one was bound. The single fragment were dissolved to the same concentration(500mM) as the cocktail solution and introduced to the crystal drop according to the same protocol, giving a protein-fragment ration of 1:280.

3.1.4 Data collection and analysis

Data collection was collected at synchrotron x-ray sources, the majority at the Diamond Light Source synchrotron in United Kingdom. Data was collected at 0-180°with small increments of 0.1-0.2°. The data was processed following a standard crystallographic pipeline. The details of the pre-processing differs between the sites but have a general procedure in common and typically uses software packages like CCP4[30], autoproc[31] and dials(Diamond Light Source). General pre-processing

procedure:

1. **Indexing and Integration**
   Indexing of reflections and integration to determine the intensity.

2. **Determination of Laue group**
   Describes the symmetry of the diffraction pattern and is used to determine space group.

3. **Scaling and Merging**
   Scales the intensities to account for physical factors and averages merged intensities in cases where the same reflection has been recorded from several angles.

4. **Determination of structure factors**
   The structure factors are determined from the intensities using different softwares like autoproc, dials, XIA2, grenADES and EDNAproc.

Processing with the purpose to create the 3D atomic model and analysis consisted of the following tasks:

- **Molecular replacement**
  Fitting of a known model to the space group of the data set to solve the phase problem. This was done with the molecular replacement module in the CCP4 suite.

- **Refinement**
  Real space fitting of the 3D atomic model to the electron density

- **Ligand fitting**
  Real space fitting of a ligand to the unoccupied electron density areas

For the visualisation, refinement and ligand fitting several task specific software's were used. The ligand fitting was performed manually with coot[32]. as well as visualisation and real-space corrections. Refinement was done with autoBUSTER[28]. All figures showing structures have been prepared in the visualisation tool pymol[33].

4

Results

4.1 Evaluation of tested crystals and crystallisation conditions

Alterations of the crystallisation conditions was done in order to minimise the protein consumption, time spent crystallising and to ensure a constant supply of well diffracting crystals. The success rate was calculated based on the percentage of crystals per well that yielded a data set suitable for model building suggested that some conditions were less productive then others. Based on this the plate design was changed to include more conditions similar to those in wells with a high success rate. The results shown in Table 4.1 suggested that conditions in the wells located in top row of the design matrix was more productive which lead to a second design. The results from the design that followed inital one are presented in Table 4.2. Comparing Design 1 and 2 one can see a slight difference in how the data is spread. The number of harvested crystals and the number of successful data collection suggests no improvement between the two designs.

Table 4.1: Results from the evaluation of the initial design. See Table 3.1 for details of the design.

Success rate	1	2	3	4	5	6
A	100%	27%	64%	67%	100%	No data
B	70%	0	No data	0	50%	No data
C	50%	0	60%	No data	0	0
D	No data	25%	No data	No data	0	No data

Table 4.2: Results after evaluating Design 2 based on the criteria stated in subsection 3.1.3.See Table 3.2 for details of the design.

Success rate	1	2	3	4	5	6
A	50%	50%	50%	44%	29%	100%
B	13%	50%	50%	60%	33%	No data
C	No data	17%	40%	33%	0	0
D	0	0	No data	No data	0	50%

From the inspection of images taken on the crystals as they sat in the loop during data collection it could be seen that the crystals that gave the best results were larger then average, was clear and were not obstructed. During the crystallisation

process it happened that crystals grew with a small layer of precipitate that could obstruct the x-rays and decrease the level of diffraction. With these two features in mind the harvesting became more selective and aimed to only target crystals with these features.
Based on the results from design 1 and 2 a third design was constructed where the ammonium sulphate concentration range were shifted lower and the pH range of Tris was narrowed. The results from design 3 are presented in Table 4.3.

Table 4.3: Design 3 shows very poor results and were quickly abandoned. See Table 3.3 for detials of the design

Success rate	1	2	3	4	5	6
A	No data	No data	No data	0	0	33%
B	No data	No data	100%	0	0	No data
C	No data	No data	100%	No data	No data	No data
D	No data	No data	No data	No data	No data	No data

As the results from both prior iterations indicated that the conditions in the top row was highly productive, design 4 was based on the conditions present in well A1-A6. As the yield improved with this design further iterations were made. The pooled result from design 4 and 5 is presented in Table 4.4.

Table 4.4: Results from design 4 and 5. The success rate was calculated with pooled data from design 4 and 5 and are presented per well, see Table 3.4 and 3.5 for details of the designs.

Success rate	1	2	3	4	5	6
A	33%	No data	No data	No data	100%	No data
B	No data	0	33%	0	50%	No data
C	No data	No data	50%	No data	No data	No data
D	0	33%	No data	50%	0	0

4.2 X-ray fragment screen

Screening of 96 cocktails resulted in 24 bound fragments. This corresponds to a hit rate of 6.6% in 361 fragment library. Out of the 96 cocktails 21 were tested prior to this project and out of those, 3 were hits. The results are included for completeness but will not be discussed in detail.
In Figure 4.1 are a surface representation of all hits presented. In Table 4.5 are the position,id and outcome of soaking experiments of the different cocktails reported as they are found in the 96 well plate. Hit(green) and no hit(red) are cases where it can be firmly determined if a fragment is bound or not. Cases with bad diffraction(yellow) refers to when the cocktail deteriorated the crystal in such way that it did not diffract. Inconclusive(orange) indicates that it could not be determined for what reason no data could be collected within the project time frame. The five

inconclusive cases were all tested once but did not yield any, or sufficiently good, data. To be able to determine if this was due to the effect of the fragments or because of mistreatment of the crystal more crystals needs to be tested. Assuming that there are 4 fragments per cocktail this means that there are 20 fragments that could be possible hits. With a hit rate of 6.6% this means that there should be 1 hit among these 20 fragments. The fact that the cocktails already are tested with a bad outcome the likelihood of a hit should be considered to be considerably lower and thus not affect the hit rate in any substantial way.

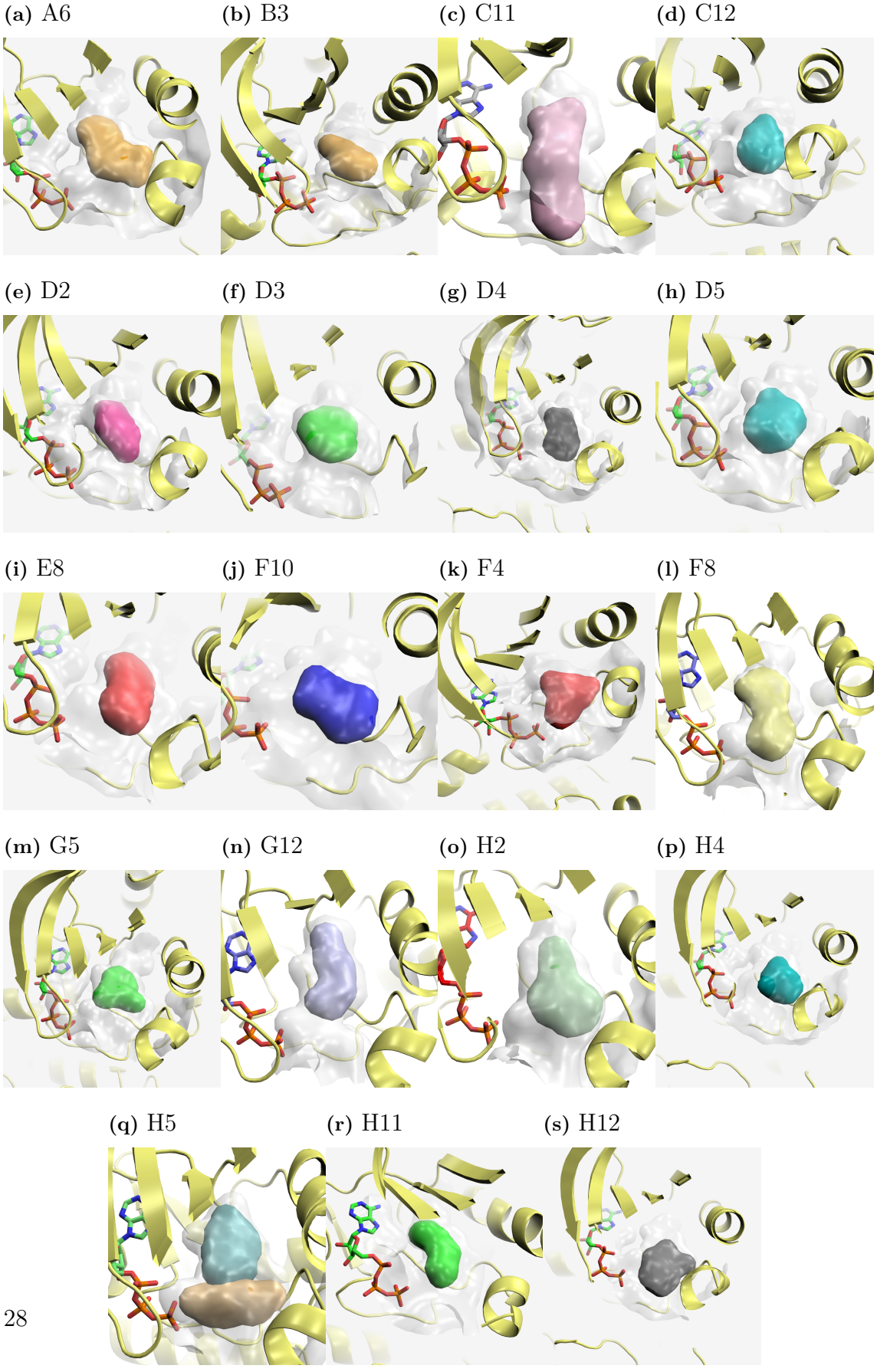
Each cocktail was introduced to several crystals, out which the two best crystals were chosen to be tested. If no data could be obtained and the reason to this appeared to be either a bad crystal or caused by the manual handling of the crystal it was re-soaked. Out of the 96 soaks performed, 25 had to be re-soaked due to bad quality data, physical damage to the loop, ice, disturbing precipitate from the soaking and crystallisation process or empty loops. From these 25, 4 additional hits was found, 7 cocktails did not yield diffraction, 9 cases with no fragments bound and 5 that remained inconclusive. The hit rate within this subset of re-soaked cocktails was 4%, indicating the value in the extra time spent.

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12
E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12

Table 4.5: Representation of the 96 cocktails and the results. Green colored cells represent a hit. Red cells no hit and yellow with red text are cocktails where the fragments are believed to cause such damage to the crystal that no diffraction occurred. Cells coloured orange are inconclusive and represents cocktails where it has not been possible, within the time frame of this project, to determine why no data could be attained.

The main focus of this screen was to identify fragments bound to the allosteric pocket of MEK1. The screen have yielded bound fragments outside this pocket and then especially in the MEK1 dimer interface. 10 cases of bound fragments have been found in this region. These cases have been refined to an extent but not so that it could be fully determined which fragment that have bound and how it is bound. To fully understand the implication of these results one needs to further refine the structures and compare the dimer interface to other constructs as well as the dimer interface between MEK1 and MEK2.

Figure 4.1: Surface representation of all fragments identified in the allosteric pocket of MEK1 during the screen. They are named after the position of the cocktail in which the binding fragment was found.



4.2.1 Crystallographic data and refinement statistics

All 24 structures used to identify the binding of a fragment have been crystallographically refined and rebuilt to ensure the correct interpretation of fragment binding. In Table 4.6 are the crystallographic and refinement statistics presented for five diverse fragments selected as they are good representatives of the distribution of where bound fragments are located and how they are bound. The same statistics for the additional 19 fragments not represented here can be found in section A.1.

Table 4.6: Statistics for refinement and data collected on crystals soaked in cocktail C11,F8,G12,H2 and H5. Numbers set within parenthesis refers to the highest resolution shell.

Cocktail	**C11**	**F8**	**G12**	**H2**	**H5**
X-ray diffraction data					
Unit cell	a=57.81	a=57.45	a=60.40	a=57.56	a=57.64
($P2_12_12_1$)[1]	b=72.07	b=71.57	b=71.64	b=71.70	b=71.92
	c=157.86	c=157.75	c=158.41	c=158.31	c=157.94
Resolution	46.6 - 2.4Å	24.06-1.78Å	48.03-2.05Å	46.55-1.74Å	46.56-1.69Å
	(2.4-2.4)Å	(1.81-1.78)Å	(2.12-2.05)Å	(1.80-1.74)Å	(1.75-1.69)Å
Completeness	99.7%	99.78%	98.4%	99.7%	99.4%
	(99.6%)	(97.68%)	(87.7%)	(97.0%)	(99.0%)
Multiplicity	6.1 (6.5)	6.47 (6.50)	6.4 (5.7)	6.3 (5.9)	6.5 (6.7)
R_{meas}[2]	0.121	0.058	0.071	0.065	0.044
	(0.875)	(1.366)	(1.838)	(1.386)	(1.479)
<I/sigI>	9.8 (2.3)	14.6 (1.3)	13.5 (0.8)	15.2 (1.1)	18.4 (1.1)
CC1/2	0.996	0.999	0.999	0.999	1.000
Refinement					
R_{fact}[3]	21.84%	21.28%	21.44%	21.42%	21.66%
R_{free}[3]	26.67%	25.57%	25.00%	23.65%	24.81%
r.m.s.[4]bond/angle	0.001Å/1.19°	0.01Å/1.12°	0.01Å/1.12°	0.01Å/1.10°	0.01Å/1.14°

[1] All crystals have the same space group

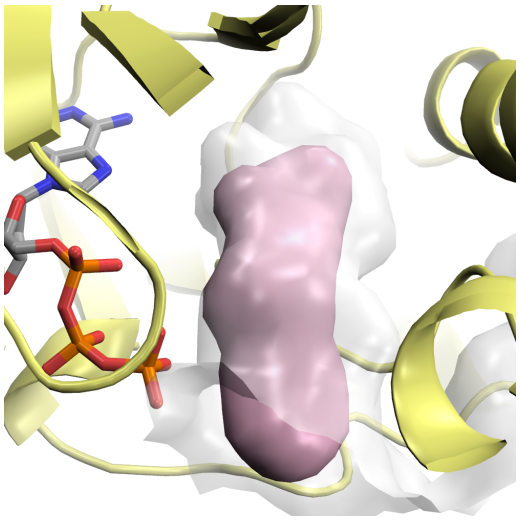
[2] $R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{j=1}^{n} |I_{hkl,j} - <I_{hkl}>|}{\sum_{hkl} \sum_{j} I_{hkl,j}}$ Redundancy independent measure of data quality

[3] $R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$ R_{fact} and R_{free} were calculated for the working set and test set respectively.

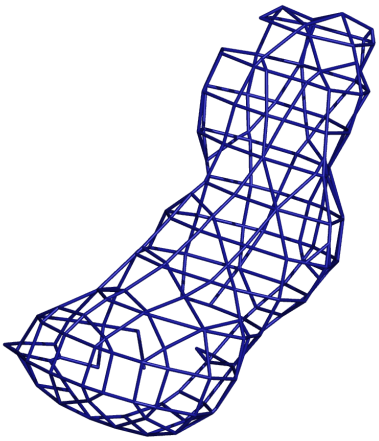
[4] r.m.s. *root mean square*

4.2.2 Fragment hits and their binding modes

The bound fragments can be classified into five different groups depending on where in the pocket they were located and what binding modes they represent. A selection of five fragments based on diversity have been chosen to provide a representative example for each group. In Figure 4.2**(a)**-4.6**(a)** the fragments are visualised as a surface representation in the context of the allosteric pocket with the bound AMP-PNP molecule as reference. In Figure 4.2**(b)**-4.6**(b)** the electron density to which the fragment was fitted are presented at 1σ. Images of all additional bound fragments and their electron densities are found in section A.2.
Figure 4.2 shows a surface representation of the bound fragment from cocktail C11. The fragment have a long shape and is the only one that extends outwards of the pocket entrance. Based on proximity analysis done in pymol[33] the fragment show interactions with Lys97, the backbone of Asp208 and a water molecule. The water molecule acts as a bridge to form interactions with Ser212 and Phe209. Both Asp208 and Phe209 are located in the back of the pocket and anchors the fragment. The same fragment is shown in Figure 4.7.



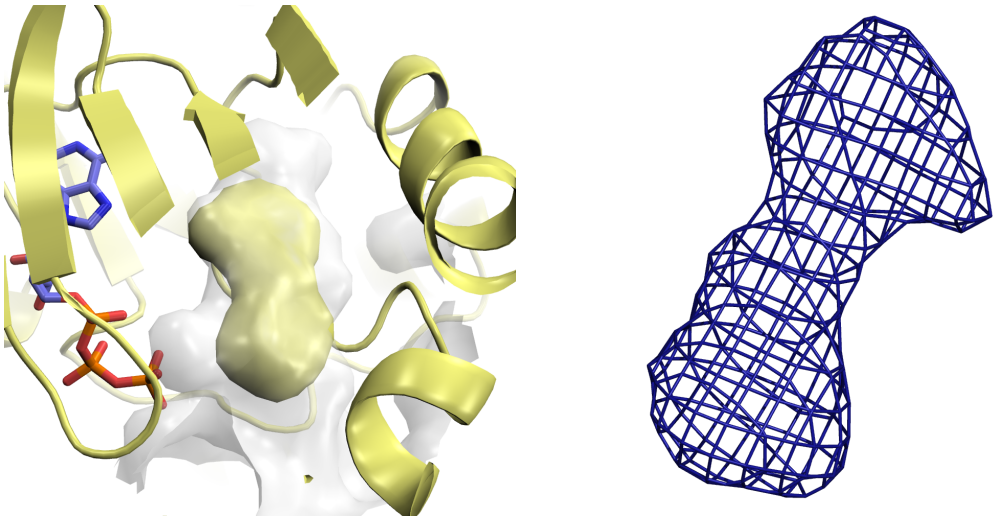
(t) Surface representation of C11



(u) Electron density at 1σ

Figure 4.2: Cocktail: C11 Surface representation of the bound fragment inside the allosteric site. The ATP binding site is seen to the left.

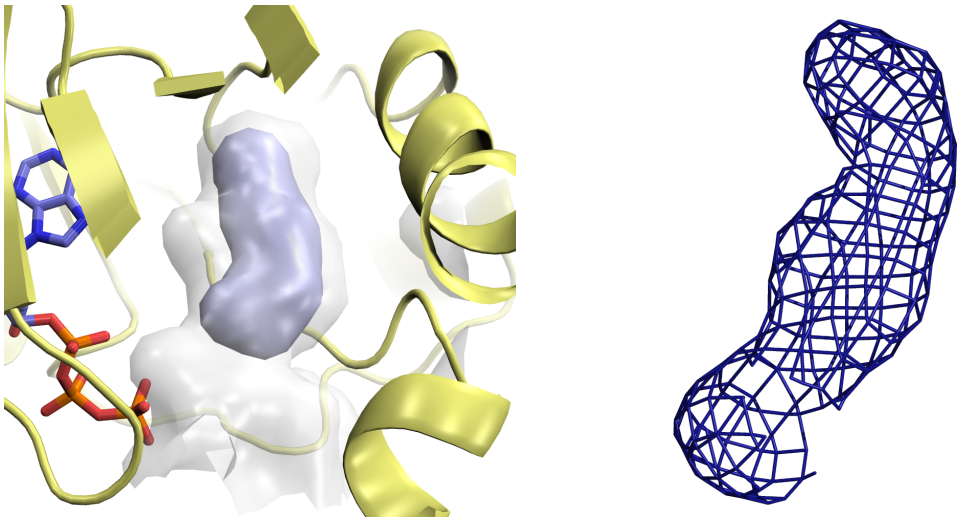
Figure 4.3 shows a fragment from cocktail F8. The central position and relativly small size represent a majority of all bound fragments. The fragment are in close proximity to Lys97 but does not interact. Instead it show interactions with Asp208 and Ser212 via a water molecule. Among the fragments found in this region the majority showed interactions with Asp208 and Phe209. The fragments in this group provide possibilities for a scaffold that can be used to link different areas of the pocket. This fragment is also represented in Figure 4.7.
In Figure 4.4 a fragment that binds deep inside the pocket is shown. The fragment was found in cocktail G12 and show interactions with Lys97, Asp208 and Phe209 along with a aromatic ring in the innermost part of the pocket.



(a) Surface representation of F8 **(b)** Electron density at 1σ

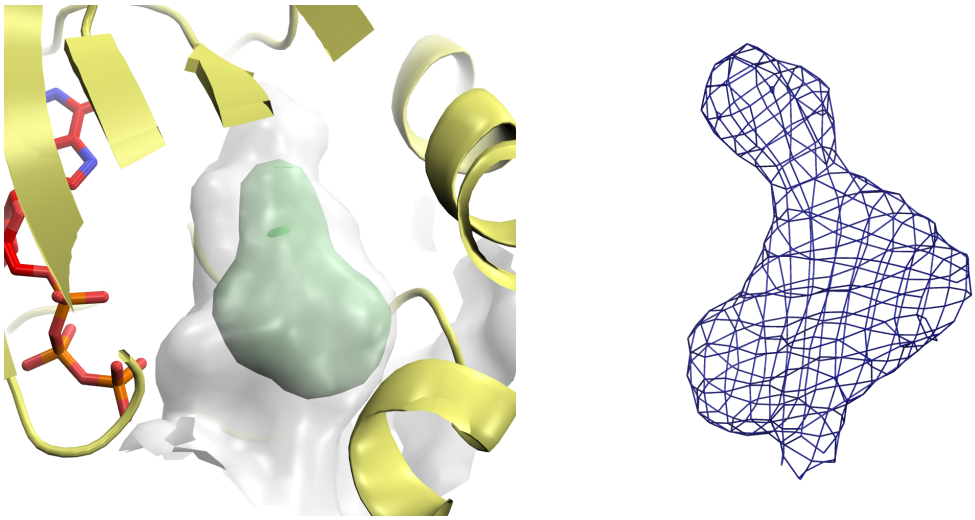
Figure 4.3: Cocktail: F8 Surface representation of the bound fragment inside the allosteric site. The ATP binding site is seen to the left.

Figure 4.5 shows a fragment from cocktail H2. This fragment has a part that is directed towards Ser212 providing a good scaffold to be utilised if the aim is to reach the Ser212 sidechains and backbone. The fragment does not show any interactions with Ser212 but is instead anchored to Phe209 and the backbone of Phe209 which in turn provides a bridge to fragments like the one described in Figure 4.4.
The surface in Figure 4.6 covers two fragments bound simultaneously. They are positioned perpendicular two each other and it is likely that the inner fragment needs to be in place in order for the outer fragment to bind. The inner most of the fragments occupy a space similar to the fragment shown in Figure 4.3. The outer most is placed so that it is interacting with both the gamma phosphate of AMP-PNP, the sidechain and backbone of Ser212 as well as Lys97. In Figure 4.7 only the outermost fragment is shown, and there with a orange surface.
In Figure 4.7 all five fragments presented above are super positioned and shown in two orientations. The top view clearly shows five distinct fragments with a large overlapping area in the center. The side view shows how the pocket has an angular shape and how the fragment from cocktail C11 follows that angle and extends towards the entrance of the pocket.



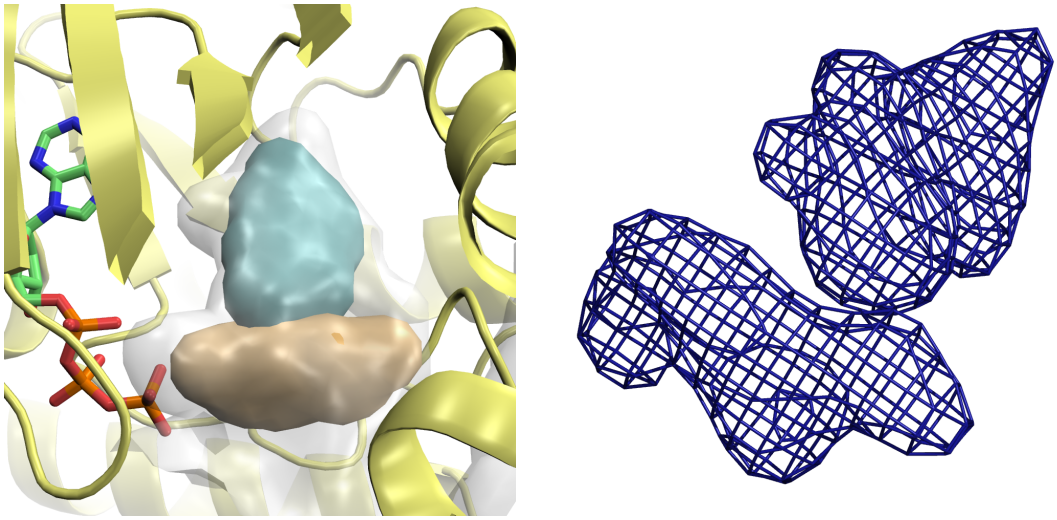
(a) Surface representation of G12 (b) Electron density at 1σ

Figure 4.4: Cocktail: G12 Surface representation of the bound fragment inside the allosteric site. The ATP binding site is seen to the left.



(a) Surface representation of H2 (b) Electron density at 1σ

Figure 4.5: Cocktail: H2 Surface representation of the bound fragment inside the allosteric site. The ATP binding site is seen to the left.



(a) Surface representation of H5 **(b)** Electron density at 1σ

Figure 4.6: Cocktail: H5 Surface representation of the bound fragment inside the allosteric site. The ATP binding site is seen to the left.

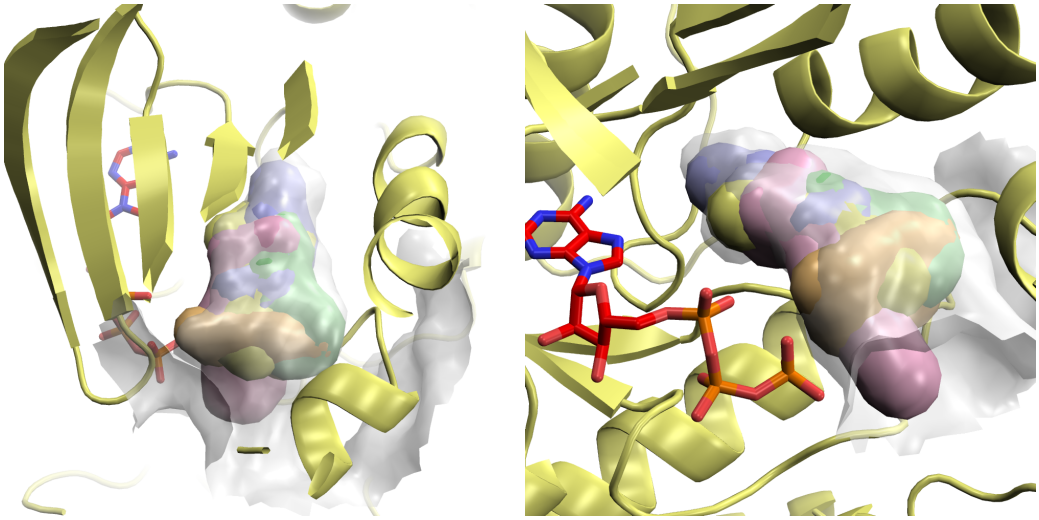


Figure 4.7: Top and side view of the superpositioned hits from cocktail C11(pink), F8(yellow), G12(blue), H2(green) and H5(orange).

5
Discussion

Crystallisation and introduction of fragments Since a crystallisation system was in place prior to the start of this project the crystal production has not been an issue which it often can be. The system that was initially used showed a good productivity and the attempts to improve on this was moderately successful. The approached used aimed to find a way to quantify the crystal quality. Unfortunately there is a large discrepancy in the amount of data available on the different designs but generally points towards the benefits if the initial designs where a larger range of ammonium sulphate and pH was used. The design where only Bis-Tris was used as a buffer did yielded a large amount of crystals with the desirable features but not enough could be tested to evaluate it properly.

The approach where the designs were evaluated based on data quality was somewhat flawed in the sense that it do not take into consideration the manual handling of the crystal or the physical state of the crystal. For example could a condition give a well diffracting crystal that was mistreated so that it lost its diffracting power thus showing as a negative result because of the manual handling and not due to the crystallisation conditions. The manual handling did improve during the project as the experimenter gained experience and training. This have most likely had the greatest affect on the productivity and make the comparisons between the different designs tricky.

The arrangement of fragments in cocktails are done in order to save time since introducing fragments one-by-one would increase the number of soaks 4-fold. This approach can be a drawback when introducing the cocktails to the mother solution if the cocktail is hard to dissolve in the particular mother solution. During the project this issue occurred in some cases. When it occurred it mostly effected the liquid dispensing which made it difficult to ensure that a sufficient amount of cocktail were transferred to the crystal. It is not possible to tell if it just one, all or a combination of fragments that causes this. If the cocktail concentration is to low in the crystal drop due to this it might result in false negatives. This is unlikely though since a binding fragment probably will dissolve and diffuse even though the others are not. The issue did not occur in any large extent within this project and the risk of false negatives is very low.

In some cases the aggregation of the cocktail was not immediate but could manifest after incubation, either as a thick skin on the droplet, to which the crystal would stick, or a skin on the crystal itself. This caused issues in a larger extent then the difficulties with liquid dispensing since this greatly affected the manual handling and caused many crystals to go to waste. This is also believed to greatly affect the diffraction. In some cases the crystal looked as if it would diffract well but showed

no result. In these cases the cocktail was assumed to deteriorate the internal order of the crystal.
The deterioration of crystals due to the soaking process were an reoccurring issue through out the project. When this occurred it manifested as completely empty data sets with small or no signs of diffraction. These cases were tested twice to ensure that it was not something else in process that might have caused this. The cause of this deterioration is very hard to find since it could be due to one or more of the fragments in the cocktail. To work around this the problematic cocktails could be convoluted and the fragments introduced one-by-one. In a best case scenario where it is assumed that it is only one fragment per cocktail that causes the deterioration and there 4 fragments per cocktail there would be 56 potential binders. With a hit rate of 6.6% that means that there could be 3 potential hits. The question to be asked is if the extra time and resources that would be needed is justified by the additional information these potentially would give.

Interpretation of electron density In a majority of the cases where there was an electron density present that suggested a bound fragment the size and shape of the density excluded all but one fragment candidate. However, in some cases the density allowed for several candidates to be fitted and it was not possible to determine which one that was bound. Alternatively, if several candidates could be bound the correct one could be identified by introducing the candidates one by one. One issue that emerged when identifying bound fragments included misinterpreted densities caused by a large solvent content occupying the pocket. In several cases determining the orientation of the bound fragment was hard. Especially when the fragment was symmetric in shape. Most often the orientation could be derived from the electron density which would show certain qualities depending on the shape and atom content of the fragment. Such a quality could be exemplified by sulfur which gives a very strong density. If the wrong orientation was assumed it can induce errors later on when trying link or extend fragments.

Bound fragments Considering how cobimetinib and the TAK-733 are positioned and bound to the pocket three features can be distinguished. The aromatic ring directed towards the inner part of the pocket, an aromatic ring directed towards Ser212 and a feature anchored in Lys97 directed outwards of the pocket with a bias towards the bound ATP and a interaction with it.
All three of these features are represented by one fragment or more, suggesting that the results of this screen have probed the pocket effectively and give a good indication of what interactions that are available. The majority of all bound fragments are positioned entirely or partly in the centre region of the pocket, giving a range of options of how to connect the three suggested features of the pocket. Looking at possible connections based on proximity, none of the fragments in the centre region show any interaction with Lys97 but are instead anchored via interactions with Asp208 and Phe209 located at the bottom of the pocket or in combination with waters acting as mediators.
Comparing fragments bound in the hydrophobic inner part of the pocket the typically aromatic ring is present as a reoccurring feature which suggests that this is

defining for this part of the pocket. Fragments positioned towards Ser212 and towards the entrance of the pocket are quite few. This is most likely because the fragment needs to be anchored in other parts of the pocket due to the relatively large entrance. Compared to the fragments bound in the centre and inner part which are generally smaller and have a more snug fit.
Taking the diversity bound fragments into consideration it is clear that the library used have been well designed and was well suited. The size of the bound fragments have an average molecular weight equal to the library average which is 194Da. The largest and smallest fragment in the library is 336Da and 110Da respectively. Among the bound fragments the largest is 255Da and the smallest 134Da. This would indicate that a majority of the fragments in the library were of a beneficial size that could probe the pocket effectively.
The next step would for example be to characterise the hits with an affinity based method like SPR to determine the affinities are and to further confirm binding. A nearest neighbour search could identify fragments with a closely related structure that have not been tested within this screen that could be of interest. In parallel is could also be investigated if some fragments could be linked or expanded to create more potent compounds. Furthermore fragments that show a suitable affinity could be tested for activity in a biochemical assay, although there is a likelihood that this would not yield any results due to the nature of allosteric inhibition.

Conclusion To conclude, x-ray crystallography as a primary screening method to identify small fragments bound to the allosteric pocket of MEK1 works very well. The results show a good diversity and could serve as good starting points for developing allosteric MEK1 inhibitors.

Bibliography

[1] JP Hughes et al. "Principles of early drug discovery". In: *British Journal of Pharmacology* 162.6 (Mar. 2011), pp. 1239–1249. ISSN: 00071188. DOI: 10.1111/j.1476-5381.2010.01127.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/21091654%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3058157%20http://doi.wiley.com/10.1111/j.1476-5381.2010.01127.x.

[2] Thierry O. Fischmann et al. "Crystal Structures of MEK1 Binary and Ternary Complexes with Nucleotides and Inhibitors". In: *Biochemistry* 48.12 (Mar. 2009), pp. 2661–2674. ISSN: 0006-2960. DOI: 10.1021/bi801898e. URL: http://www.ncbi.nlm.nih.gov/pubmed/19161339%20http://pubs.acs.org/doi/abs/10.1021/bi801898e.

[3] G. Pearson et al. "Mitogen-Activated Protein (MAP) Kinase Pathways: Regulation and Physiological Functions". In: *Endocrine Reviews* 22.2 (Apr. 2001), pp. 153–183. ISSN: 0163-769X. DOI: 10.1210/er.22.2.153. URL: http://press.endocrine.org/doi/10.1210/edrv.22.2.0428.

[4] Jeffrey F Ohren et al. "Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition". In: *Nature Structural & Molecular Biology* 11.12 (Dec. 2004), pp. 1192–1197. ISSN: 1545-9993. DOI: 10.1038/nsmb859. URL: http://www.ncbi.nlm.nih.gov/pubmed/15543157%20http://www.nature.com/articles/nsmb859.

[5] D W Hommes, M P Peppelenbosch, and S J H van Deventer. "Mitogen activated protein (MAP) kinase signal transduction pathways and novel anti-inflammatory targets." In: *Gut* 52.1 (Jan. 2003), pp. 144–51. ISSN: 0017-5749. URL: http://www.ncbi.nlm.nih.gov/pubmed/12477778%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1773518.

[6] P J Roberts and C J Der. "Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer". In: *Oncogene* 26.22 (May 2007), pp. 3291–3310. ISSN: 0950-9232. DOI: 10.1038/sj.onc.1210422. URL: http://www.nature.com/articles/1210422.

[7] Judith S Sebolt-Leopold. "Development of anticancer drugs targeting the MAP kinase pathway". In: *Oncogene* 19.56 (Dec. 2000), pp. 6594–6599. ISSN: 0950-9232. DOI: 10.1038/sj.onc.1204083. URL: http://www.nature.com/articles/1204083.

[8] G. Hatzivassiliou et al. "Mechanism of MEK inhibition determines efficacy in mutant KRAS- versus BRAF-driven cancers." In: *Nature* 501 (2013), pp. 232–236. DOI: 10.2210/PDB4LMN/PDB. URL: https://www.rcsb.org/structure/4lmn.

[9] Q. Dong et al. “Discovery of TAK-733, a potent and selective MEK allosteric site inhibitor for the treatment of cancer.” In: *Bioorg.Med.Chem.Lett.* 21 (2011), pp. 1315–1319. DOI: 10.2210/PDB3PP1/PDB. URL: https://www.rcsb.org/structure/3PP1.

[10] Robert Roskoski. “Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes”. In: *Pharmacological Research* 103 (Jan. 2016), pp. 26–48. ISSN: 1043-6618. DOI: 10.1016/J.PHRS.2015.10.021. URL: https://www.sciencedirect.com/science/article/pii/S1043661815301298?via%3Dihub.

[11] Peng Wu, Mads H. Clausen, and Thomas E. Nielsen. “Allosteric small-molecule kinase inhibitors”. In: *Pharmacology & Therapeutics* 156 (Dec. 2015), pp. 59–68. ISSN: 0163-7258. DOI: 10.1016/J.PHARMTHERA.2015.10.002. URL: https://www.sciencedirect.com/science/article/pii/S0163725815001837.

[12] Norbert Berndt, Rezaul M Karim, and Ernst Schönbrunn. “Advances of small molecule targeting of kinases”. In: *Current Opinion in Chemical Biology* 39 (Aug. 2017), pp. 126–132. ISSN: 1367-5931. DOI: 10.1016/J.CBPA.2017.06.015. URL: https://www.sciencedirect.com/science/article/pii/S1367593117300911.

[13] Aron W Fenton. “Allostery: an illustrated definition for the ’second secret of life’.” In: *Trends in biochemical sciences* 33.9 (Sept. 2008), pp. 420–5. ISSN: 0968-0004. DOI: 10.1016/j.tibs.2008.05.009. URL: http://www.ncbi.nlm.nih.gov/pubmed/18706817%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2574622.

[14] Robert Roskoski. “Allosteric MEK1/2 inhibitors including cobimetanib and trametinib in the treatment of cutaneous melanomas”. In: *Pharmacological Research* 117 (Mar. 2017), pp. 20–31. ISSN: 1043-6618. DOI: 10.1016/J.PHRS.2016.12.009. URL: https://www.sciencedirect.com/science/article/pii/S1043661816313093.

[15] Rob L M Van Montfort and Paul Workman. “Structure-based drug design: aiming for a perfect fit”. In: *Essays in Biochemistry* 61 (2017), pp. 431–437. ISSN: 0071-1365. DOI: 10.1042/EBC20170052. URL: https://doi.org/10.1042/EBC20170052.

[16] Disha Patel, Joseph D. Bauman, and Eddy Arnold. “Advantages of crystallographic fragment screening: Functional and mechanistic insights from a powerful platform for efficient drug discovery”. In: *Progress in Biophysics and Molecular Biology* 116.2-3 (2014), pp. 92–100. ISSN: 00796107. DOI: 10.1016/j.pbiomolbio.2014.08.004. URL: http://dx.doi.org/10.1016/j.pbiomolbio.2014.08.004.

[17] György M. Keserű et al. “Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia”. In: *Journal of Medicinal Chemistry* 59.18 (Sept. 2016), pp. 8189–8206. ISSN: 0022-2623. DOI: 10.1021/acs.jmedchem.6b00197. URL: http://pubs.acs.org/doi/10.1021/acs.jmedchem.6b00197.

[18] Alexander McPherson and Jose A. Gavira. “Introduction to protein crystallization”. In: *Acta Crystallographica Section F Structural Biology Com-*

munications 70.1 (Jan. 2014), pp. 2–20. ISSN: 2053-230X. DOI: 10.1107/S2053230X13033141. URL: http://www.ncbi.nlm.nih.gov/pubmed/24419610%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3943105%20http://scripts.iucr.org/cgi-bin/paper?S2053230X13033141.

[19] Michael J. Aziz, David Turnbull, and Morton B. Panish. "Crystal growth". In: *Access Science* (2014). DOI: 10.1036/1097-8542.171500. URL: https://www-accessscience-com.proxy.lib.chalmers.se/content/crystal-growth/171500.

[20] David H. Klein and Louis Gordon. "Nucleation". In: *Access Science* (2014). DOI: 10.1036/1097-8542.460500. URL: https://www-accessscience-com.proxy.lib.chalmers.se/content/nucleation/460500.

[21] Gale. Rhodes. *Crystallography made crystal clear : a guide for users of macromolecular models*. Complementary Science. Academic Press, Mar. 1993, p. 202. ISBN: 9780125870757.

[22] Jan Drenth. *Principles of Protein X-ray Crystallography*. Springer Advanced Texts in Chemistry. New York, NY: Springer New York, Jan. 1999. ISBN: 978-1-4757-3094-4. DOI: 10.1007/978-1-4757-3092-0. URL: http://link.springer.com/10.1007/978-1-4757-3092-0.

[23] Terese M. Bergfors. *Protein crystallization: Techniques, Strategies, and Tips: A Laboratory Manual*. La Jolla, CA: International University Line, 1999. ISBN: 0-9636817-5-3.

[24] Wikimedia Commons. *File:X-ray diffraction pattern 3clpro.jpg — Wikimedia Commons{,} the free media repository*. 2016. URL: https://commons.wikimedia.org/w/index.php?title=File:X-ray_diffraction_pattern_3clpro.jpg&oldid=210910759.

[25] M M Woolfson. "The development of structural x-ray crystallography". In: *Physica Scripta* 93.3 (Mar. 2018), p. 032501. ISSN: 0031-8949. DOI: 10.1088/1402-4896/aa9c30. URL: http://stacks.iop.org/1402-4896/93/i=3/a=032501?key=crossref.83af955ce12274eb628f82a75abe53a7.

[26] Lawrence F. Dahl. "X-ray crystallography". In: *Access Science* (2014). DOI: 10.1036/1097-8542.750500. URL: https://www-accessscience-com.proxy.lib.chalmers.se/content/x-ray-crystallography/750500.

[27] E Prince and D. M. Collins. "Other refinement methods". In: *International Tables for Crystallography*. Chester, England: International Union of Crystallography, Oct. 2006, pp. 689–692. DOI: 10.1107/97809553602060000610. URL: http://xrpp.iucr.org/cgi-bin/itr?url_ver=Z39.88-2003&rft_dat=what%3Dchapter%26volid%3DCb%26chnumo%3D8o2%26chvers%3Dv0001.

[28] Oliver S Smart et al. "Exploiting structure similarity in refinement: automated NCS and target-structure restraints in {\it BUSTER}". In: *Acta Crystallographica Section D* 68.4 (Apr. 2012), pp. 368–380. DOI: 10.1107/S0907444911056058. URL: https://doi.org/10.1107/S0907444911056058.

[29] Louise Barlind. *Personal communication*. Gothenburg, 2018.

[30] Martyn D. Winn et al. "Overview of the CCP 4 suite and current developments". In: *Acta Crystallographica Section D Biological Crystallography* 67.4 (Apr. 2011), pp. 235–242. ISSN: 0907-4449. DOI: 10.1107/

S0907444910045749. URL: http://scripts.iucr.org/cgi-bin/paper?S0907444910045749.

[31] Clemens Vonrhein et al. “Data processing and analysis with the autoPROC toolbox”. In: *Acta Crystallographica Section D Biological Crystallography* 67.4 (Apr. 2011), pp. 293–302. ISSN: 0907-4449. DOI: 10.1107/S0907444911007773. URL: http://scripts.iucr.org/cgi-bin/paper?S0907444911007773.

[32] Paul Emsley et al. “Features and Development of Coot”. In: *Acta Crystallographica Section D - Biological Crystallography* 66 (2010), p. 486.

[33] LLC Schrödinger. “The {JyMOL} Molecular Graphics Development Component, Version˜1.8”. Nov. 2015.

A

Appendix

A.1 Crystallographic data and refinement statistics

Table A.1: Statistics for refinement and data collected on crystals soaked in cocktail A6,B3,F4,C12 and D2. Numbers set within parenthesis refers to the highest resolution shell.

Cocktail	**A6**	**B3**	**F4**	**C12**	**D2**
X-ray diffraction data					
Unit cell	a=57.88	a=57.27	a=57.74	a=58.2	a=57.74
($P2_12_12_1$)[1]	b=72.03	b=71.99	b=71.68	b=72.4	b=71.14
	c=157.78	c=158.10	c=157.78	c=158.4	c=158.18
Resolution	53.19 - 1.73Å	100-1.9Å	54.24-1.78Å	65.8-1.8Å	46.64-2.30Å
	(1.76-1.73)Å	(2.0-1.9)Å	(1.81-1.78)Å	(1.9-1.8)Å	(2.36-2.30)Å
Completeness	99.96%	99.4%	99.67%	92.7%	99.8%
	(99.88%)	(97.0%)	(98.24%)	(100.0%)	(97.9%)
Multiplicity	6.4 (5.7)	6.3 (5.9)	6.49 (6.58)	6.2 (6.5)	6.6 (6.7)
R_{meas}[2]	0.086	0.066	0.057	0.88	0.090
	(1.84)	(1.703)	(1.301)	(0.983)	(1.678)
<I/sigI>	13.9 (1.2)	14.7 (1.1)	14.5 (1.3)	13.4 (2.0)	10.9 (1.3)
CC1/2	0.999	0.999	0.996	0.997	0.998
Refinement					
R_{fact}[3]	21.7%	21.5%	21.12%	21.37%	21.94%
R_{free}[3]	24.8%	25.2%	25.18%	24.34%	28.10%
r.m.s.[4]bond/angle	0.01Å/1.08°	0.001Å/1.16°	0.01Å/1.096°	0.01Å/1.14°	0.01Å/1.16°

[1] All crystals have the same space group

[2] $R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}|I_{hkl,j}-<I_{hkl}>|}{\sum_{hkl}\sum_{j} I_{hkl,j}}$ Redundancy independent measure of data quality

[3] $R = \frac{\sum||F_{obs}|-|F_{calc}||}{\sum|F_{obs}|}$ R_{fact} and R_{free} were calculated for the working set and test set respectively.

[4] r.m.s. *root mean square*

Table A.2: Statistics for refinement and data collected on crystals soaked in cocktail D3,D4,D5,E8 and F10.Numbers set within parenthesis refers to the highest resolution shell.

Cocktail	**D3**	**D4**	**D5**	**E8**	**F10**
X-ray diffraction data					
Unit cell	a=57.26	a=57.44	a=57.29	a=57.66	a=57.6
($P2_12_12_1$)[1]	b=71.53	b=71.86	b=71.67	b=71.6	b=71.4
	c=157.46	c=157.63	c=158.23	c=157.41	c=158.4
Resolution	46.31-2.46Å	46.42-2.04Å	46.4-1.99Å	46.52-2.0Å	9.6-2.20Å
	(2.52-2.46)Å	(2.09-2.04)Å	(2.04-1.99)Å	(2.05-2.0)Å	(2.2-2.2)Å
Completeness	99.6%	99.6%	99.7%	98.6%	99.4%
	(94.9%)	(95.0%)	(97.3%)	(96.8%)	(93.4%)
Multiplicity	6.6 (6.5)	6.5 (6.0)	6.7 (6.7)	5.6 (5.2)	6.3 (5.3)
R_{meas}[2]	0.097	0.071	0.051	0.075	0.159
	(1.17)	(1.197)	(0.935)	(1.027)	(2.670)
<I/sigI>	11.1 (1.4)	13.9 (2.6)	17.6 (2.3)	13.0 (1.4)	9.02 (0.63)
CC1/2	0.998	0.998	0.999	0.999	0.998
Refinement					
R_{fact}[3]	20.44%	20.98%	20.77%	20.38%	19.33%
R_{free}[3]	27.99%	25.40%	24.19%	23.77%	25.22%
r.m.s.[4]bond/angle	0.01Å/1.20°	0.001Å/1.14°	0.01Å/1.13°	0.001Å/1.16°	0.01Å/1.22°

[1] All crystals have the same space group

[2] $R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}|I_{hkl,j}-<I_{hkl}>|}{\sum_{hkl}\sum_{j}I_{hkl,j}}$ Redundancy independent measure of data quality

[3] $R = \frac{\sum||F_{obs}|-|F_{calc}||}{\sum|F_{obs}|}$ R_{fact} and R_{free} were calculated for the working set and test set respectively.

[4] r.m.s. *root mean square*

Table A.3: Statistics for refinement and data collected on crystals soaked in cocktail H12,G5,H4, and H11.Numbers set within parenthesis refers to the highest resolution shell.

Data set id:	**H12**	**G5**	**H4**	**H11**
X-ray diffraction data				
Unit cell	a=57.77	a=57.94	a=57.72	a=60.03
($P2_12_12_1$)[1]	b=72.26	b=72.61	b=71.89	b=71.45
	c=158.22	c=158.03	c=157.73	c=158.41
Resolution	27.13 - 1.74Å	42.64-1.67Å	46.58-1.88Å	32.57-2.24Å
	(1.77-1.74)Å	(1.70-1.67)Å	(1.99-1.88)Å	(2.28-2.24)Å
Completeness	99.85%	99.41%	99.2%	99.84%
	(99.29%)	(97.0%)	(96.94%)	(99.7%)
Multiplicity	6.45 (5.79)	6.26 (4.3)	6.25 (5.3)	6.24 (6.44)
R_{meas}[2]	0.050	0.055	0.132	0.145
	(1.626)	(1.393)	(2.178)	(1.725)
<I/sigI>	16.5 (1.0)	16.1 (1.0)	9.82 (0.75)	8.0 (1.1)
CC1/2	0.999	0.999	0.998	0.996
Refinement				
R_{fact}[3]	21.24%	21.91%	21.45%	20.84%
R_{free}[3]	23.74%	23.46%	26.69%	25.80%
r.m.s.[4]bond/angle	0.01Å/1.10°	0.01Å/1.07°	0.001Å/1.20°	0.001Å/1.15°

[1] All crystals have the same space group

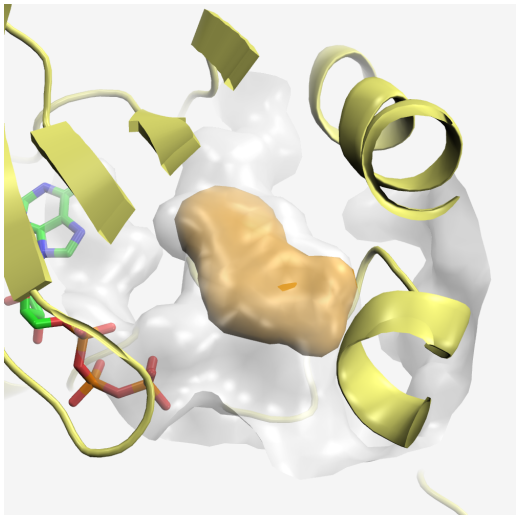
[2] $R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_{j=1}^{n}|I_{hkl,j}-<I_{hkl}>|}{\sum_{hkl}\sum_{j}I_{hkl,j}}$ Redundancy independent measure of data quality

[3] $R = \frac{\sum||F_{obs}|-|F_{calc}||}{\sum|F_{obs}|}$ R_{fact} and R_{free} were calculated for the working set and test set respectively.

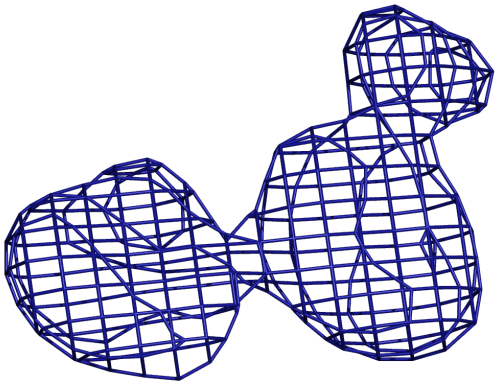
[4] r.m.s. *root mean square*

A.2 Density gallery

Below are all hits represented with a surface next to the electron density map to which they have been fitted.

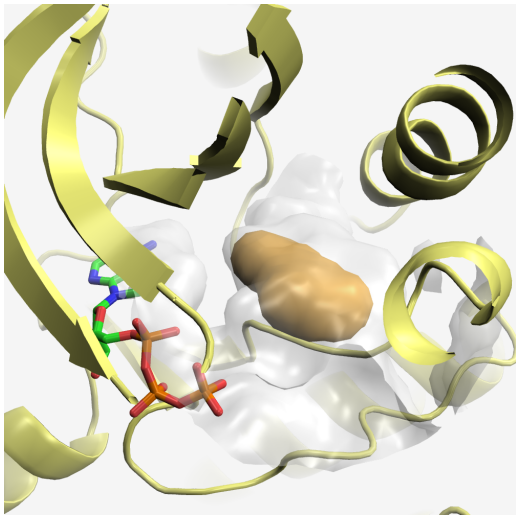


(a) Surface representation of A6

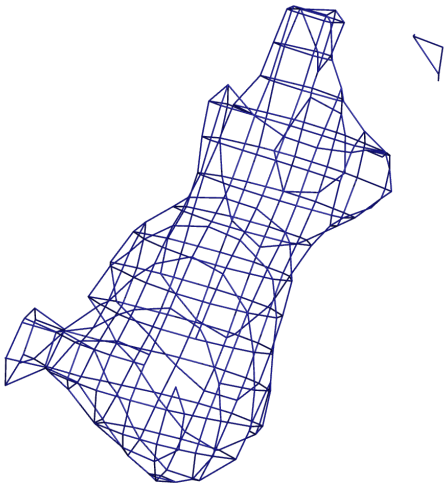


(b) Electron density at 1σ

Figure A.1: Cocktail: A6

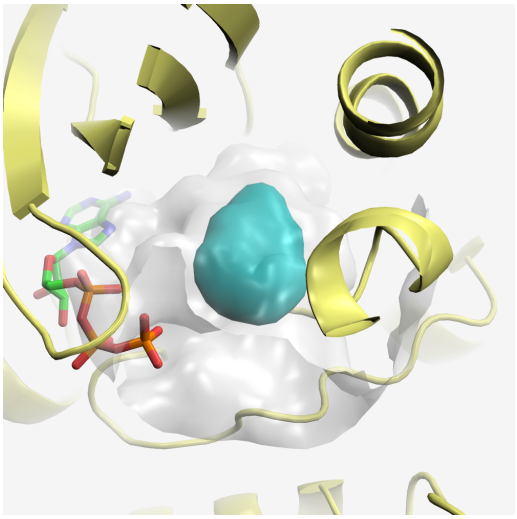


(a) Surface representation of B3

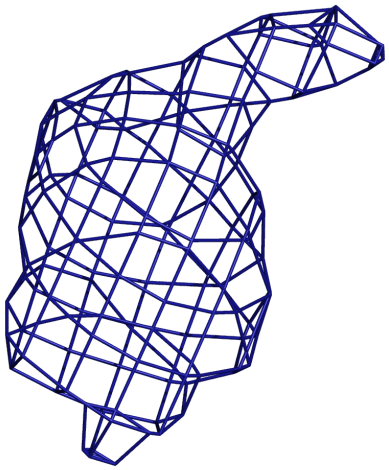


(b) Electron density at 1σ

Figure A.2: Cocktail: B3

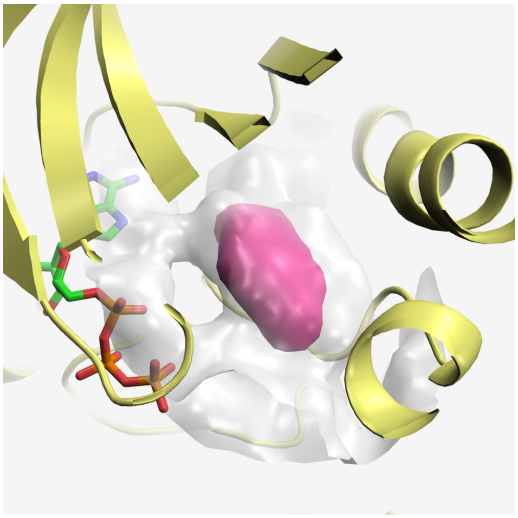


(a) Surface representation of C12

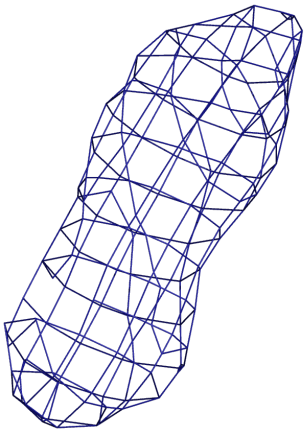


(b) Electron density at 1σ

Figure A.3: Cocktail: C12

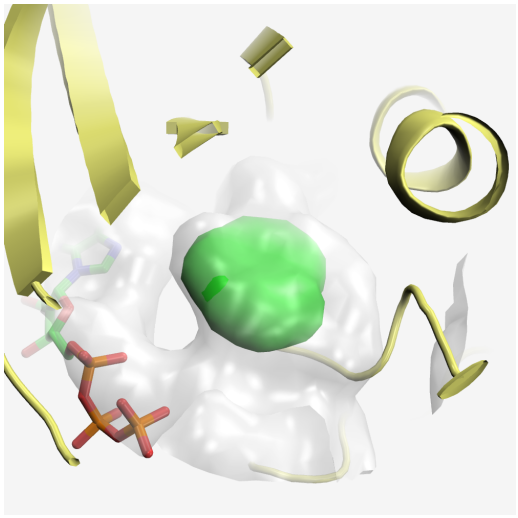


(a) Surface representation of D2

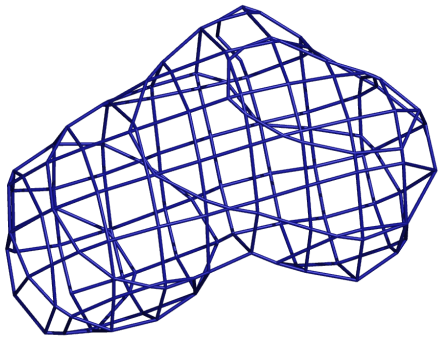


(b) Electron density at 1σ

Figure A.4: Cocktail: D2

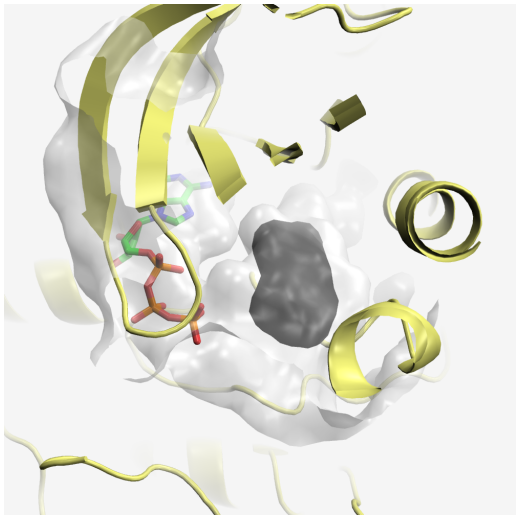


(a) Surface representation of D3

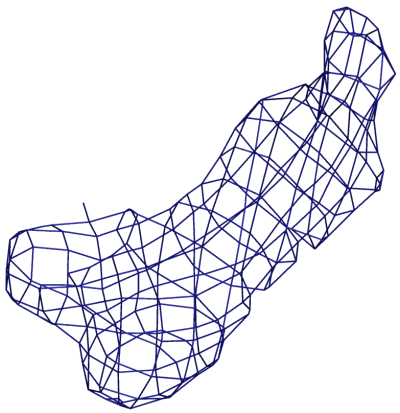


(b) Electron density at 1σ

Figure A.5: Cocktail: D3

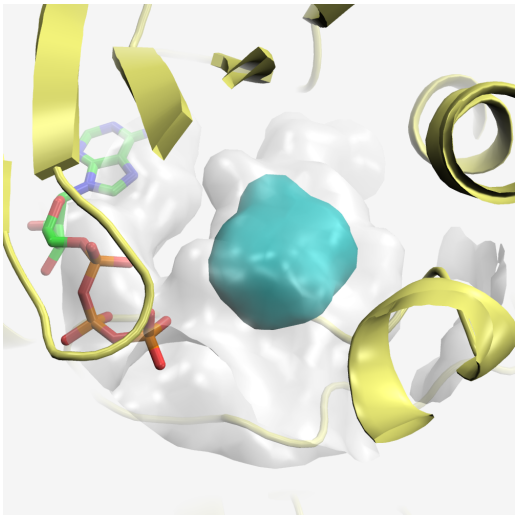


(a) Surface representation of D4

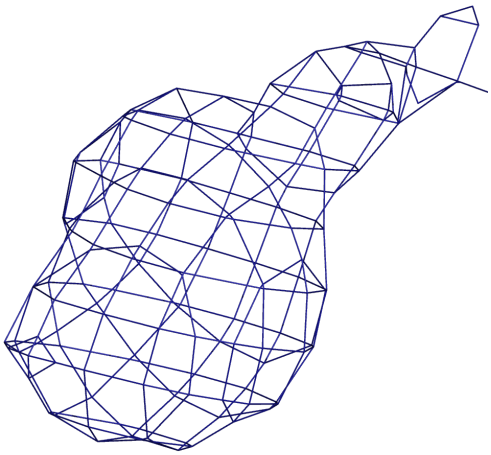


(b) Electron density at 1σ

Figure A.6: Cocktail: D4

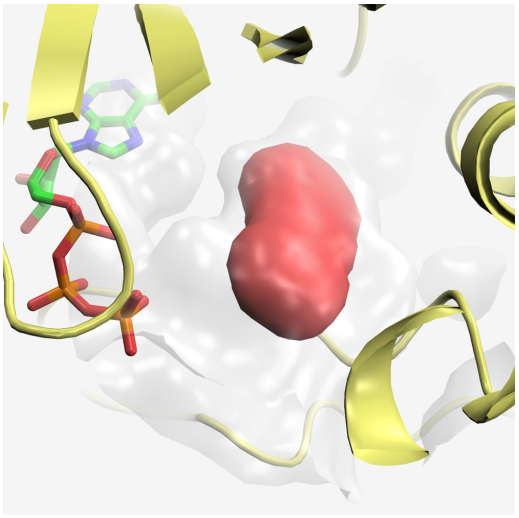


(a) Surface representation of D5

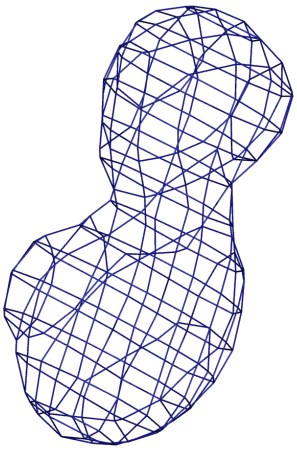


(b) Electron density at 1σ

Figure A.7: Cocktail: D5

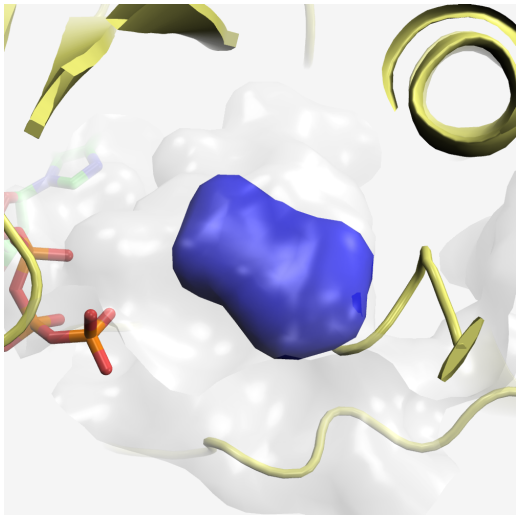


(a) Surface representation of E8

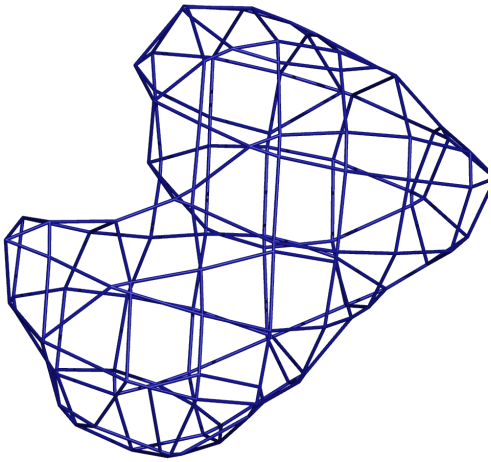


(b) Electron density at 1σ

Figure A.8: Cocktail: E8

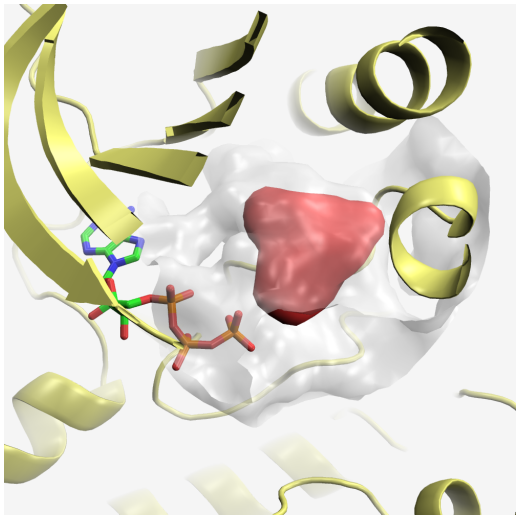


(a) Surface representation of F10

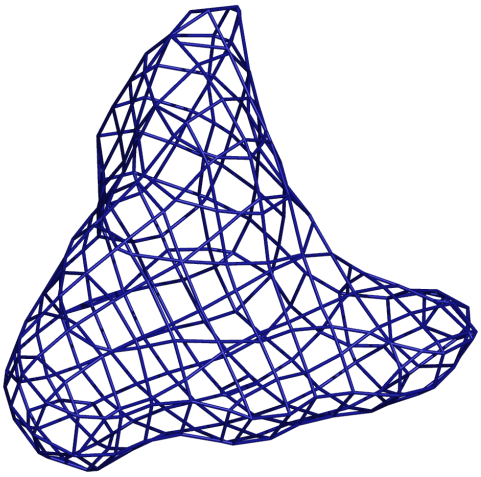


(b) Electron density at 1σ

Figure A.9: Cocktail: F10

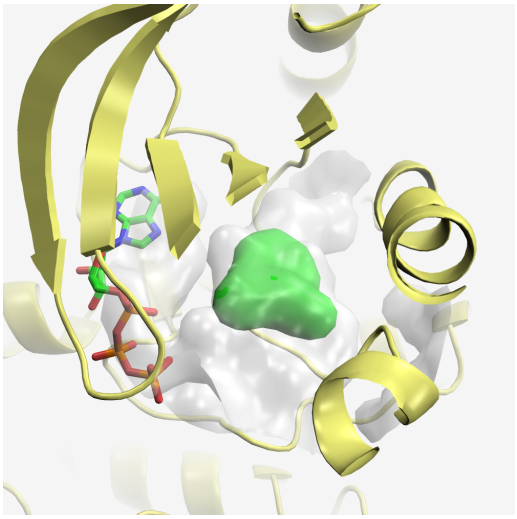


(a) Surface representation of F4

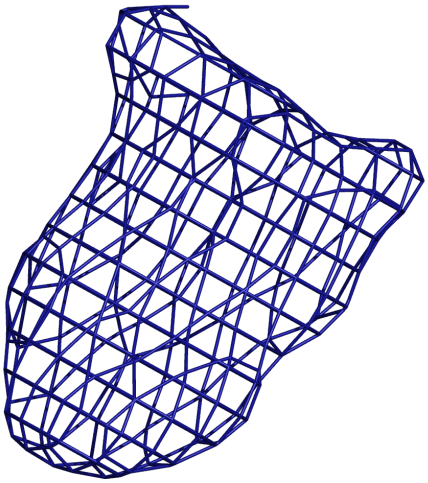


(b) Electron density at 1σ

Figure A.10: Cocktail: F4

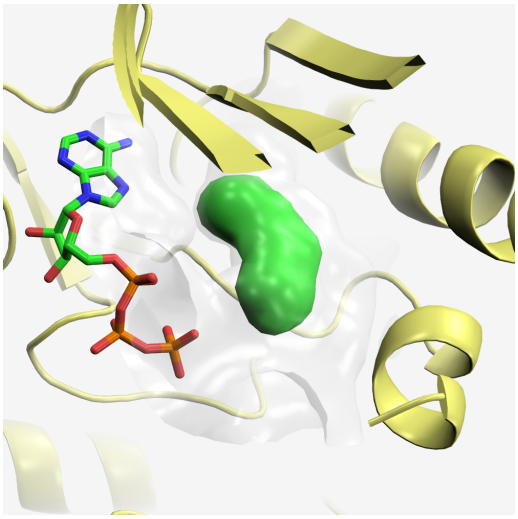


(a) Surface representation of G5

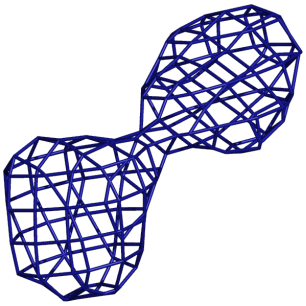


(b) Electron density at 1σ

Figure A.11: Cocktail: G5

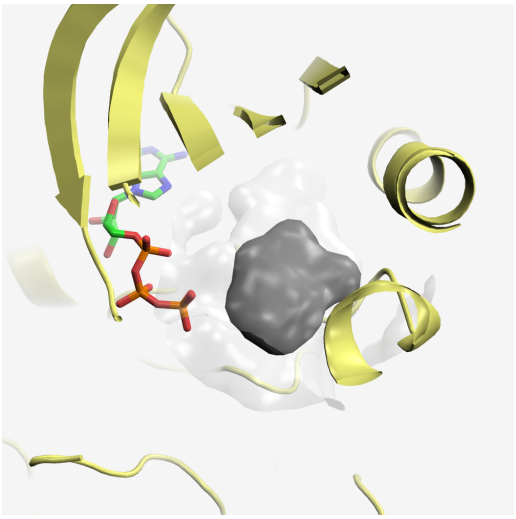


(a) Surface representation of H11

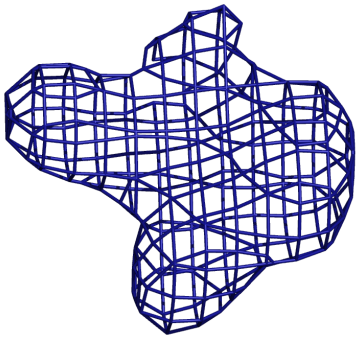


(b) Electron density at 1σ

Figure A.12: Cocktail: H11

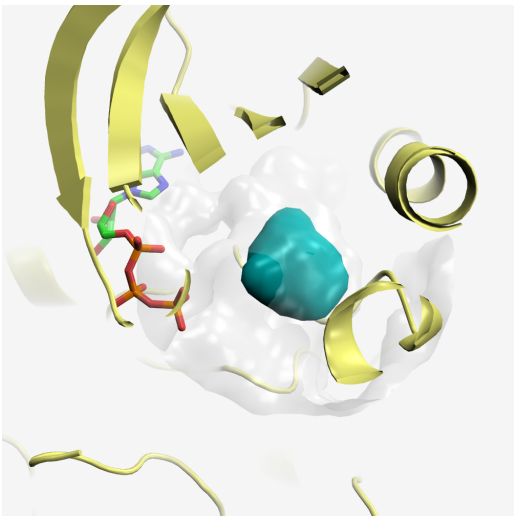


(a) Surface representation of H12

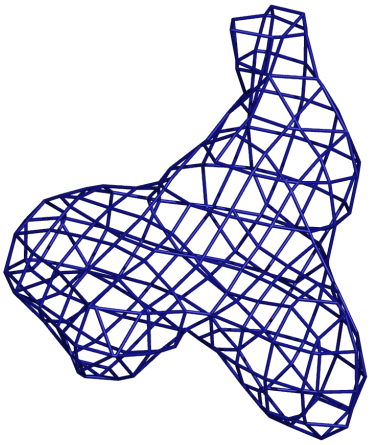


(b) Electron density at 1σ

Figure A.13: Cocktail: H12



(a) Surface representation of H4



(b) Electron density at 1σ

Figure A.14: Cocktail: H4

Figure A.15: How a general crystal could look like during data collection

A.3 Representative example of a tested crystal