Structural analysis of protein-small molecule interactions by a crystallographic and spectroscopic approach

by

Robert Fagiewicz

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE IN ADVANCED SPECTROSCOPY IN CHEMISTRY

Dipartimento di Chimica Industriale “Toso Montanari”
Dipartimento di Chimica “Giacomo Ciamician”

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA
Bologna, Italy

2017

Supervisor

Simona Fermani

Co-supervisors

Mirko Zaffagnini

Damiano Genovese

Abstract

Modern spectroscopic techniques grant various methods for a protein structure determination among with a ligand interaction. This work aims at probing the structural insights of a protein-small molecule interaction with biocrystallography and optical spectroscopies. Two independent systems were investigated in frame of this thesis. The first one involves flavoenzyme interaction with a natural nucleotide as a cofactor required for its catalytic activity and work was purely based on macromolecular crystallography. The second concerns incorporation of a synthetic fluorescent ligand into a model protein as a solution for hydrophobicity of the probe. Due to the nature of the probe optical spectroscopies (such as absorption, fluorescence lifetime, circular dichroism) were effectively employed together with crystallographic methodology.

Abbreviations

APX – ascorbate peroxidase

BODIPY – boron-dipyrromethene (also referred as pyrromethene597)

BSA – bovine serum albumin

DHAR – dehydroascorbate reductase

ESA – equine serum albumin

FAD – flavin adenine dinucleotide

$FADH^-$ – flavin adenine dinucleotide (reduced)

GR1 – glutathione reductase (cytoplasm isoform)

GR2 – glutathione reductase (chloroplast isoform)

GSH – glutathione (reduced)

GSR – glutathione disulfide reductase

GSSG – glutathione disulfide (oxidized)

HSA – human serum albumin

MDAR – monodehydroascorbate reductase

NAD^+ – nicotinamide adenine dinucleotide

$NADP^+$ – nicotinamide adenine dinucleotide phosphate

NADPH – nicotinamide adenine dinucleotide phosphate (reduced)

PEG – polyethylene glycol

ROS – reactive oxygen species

TCEP – tris(2-carboxyethyl)phosphine hydrochloride

Table of contents

PROJECT2

List of Figures

List of Tables

Acknowledgements

I would first thank my thesis supervisor Prof. Simona Fermani from the Department of Chemistry of the University of Bologna. It would be though to list the number of times when I found help in her office. At the moment I cannot think of a person that would devote this amount of time for many enlightening conversations, explanations, and discussions in regards to this project. After working in Prof. Fermani's group I am sure I want to continue my adventure with structural biology and I wish everyone to start scientific research with people like her. Not only a knowledgeable and highly skilled scientist but also an understanding tutor, guide, and a friend. I truly hope we will be able to cooperate during my future career.

I would also like to thank my co-supervisors: Mirko Zaffagnini (Department of Pharmacy and Biotechnology) and Damiano Genovese (Department of Chemistry). Without their enthusiastic input, many issues would remain unsolved. I am very grateful for a number of valuable comments on this thesis.

From people that are more related to my life during this degree rather than the project itself, the list is long and impossible to describe their contribution. Nevertheless they played an important role in the everyday work, study/lunch/party time, office work and more generally - my life during ASC: coordinator of the ASC network – Prof. Sylvain Cristol, local coordinators in Lille and Bologna – Prof. Cedric Lion and Prof. Marco Giorgetti and all the ASC network teachers for this amazing program, administrative officers – Francine Chanier and Aude Zeglin for helping out with all that painful paperwork, my amazing friends for the support in tough moments, and last but not least all the people from Biocrystallization and Biomineralization group at University of Bologna – Prof. Giuseppe Falini, Dr. Matteo Di Giosia, Dr. Michela Reggi, Devis Montroni and Giulia Magnabosco for making the lab work more enjoyable than ever and for hundreds of conversations (those scientific and those far away from science).

Chapter 1 - Preamble

I. Structural studies of biological macromolecules – biocrystallography

Contemporary structural biology employs various powerful techniques to investigate the structure of macromolecules at different levels and resolutions; among them X-ray or neutron diffraction (XRD), Small Angle X-ray Scattering (SAXS), Nuclear Magnetic Resonance (NMR), and Cryo-Electron Microscopy (Cryo-EM). All of them plays at the edge of physics, biology, chemistry, and engineering. The field requires very often hybrid approach, thus combining two or more techniques. Biocrystallography is the experimental science aimed to the study of the 3D structure of biological macromolecules at atomic resolution. It is based on the analysis of a diffraction pattern of a crystal, targeted by a radiation beam, in general X-ray. Although the crystallization of biological macromolecules could be the bottleneck of the entire procedure, biocrystallography remains dominant and allows for an atomic resolution structure, eventually giving access to the function evaluation.

Historically this field is not only knowledge-based but also technology and methodology driven. Regular advancements had been done in emerging aspects of biocrystallography, starting from protein expression and purification and ending up with diffraction data collection at synchrotron sources and model refinement. As Giegé *et al*. (2010) nicely phrased this evolution:

"Over the years with the increasing number of solved crystal structures, biocrystallography reached the mature age and transformed into structural biology."

Difficult to argue with such statement considering the number of cooperative networks doing extensive work in the structure determination by employing more and more advanced technologies. Many European institutes allow researchers to access almost any kind of instrumentation with a support of experienced scientists (i.e. EMBL, IBS, Instruct facilities).

Figure 1 represents steps that are essential in 3D structure determination by the crystallographic approach. The very first step is the preparation of the target macromolecule sample, that is based on biological methodologies (cloning, expression, purification and purity assessment). The further three steps deal with rather

interdisciplinary approaches to get high-quality diffraction data from the protein crystal. The main condition to reach this aim is to grow highly ordered single crystals. Methods were developed to screen initial conditions and pick the right method for crystal growth. The complexity of the crystallization matrix makes it more a trial and error approach, but when the first crystals are obtained an optimization procedure can be applied by rationally varying few parameters with the aim of improving morphology and dimensions. Sometimes these two steps can require a long time. The determination of the final 3D structure consists of two stages: the phase determination which is not a trivial task in the case of biological macromolecules and the structure refinement. Many software are currently available for the diffraction data analysis, phase determination and structure refinement. Being constantly updated they keep the crystallographers with most recent advancements.

A crystallographic project can be stuck at any of these stages forcing search for an alternative solutions and re-attempting the job all over again. As will be demonstrated in this work, it is the case where some problems are difficult to overcome but remains solvable to some extent. Methodology details will be provided further on in this work.

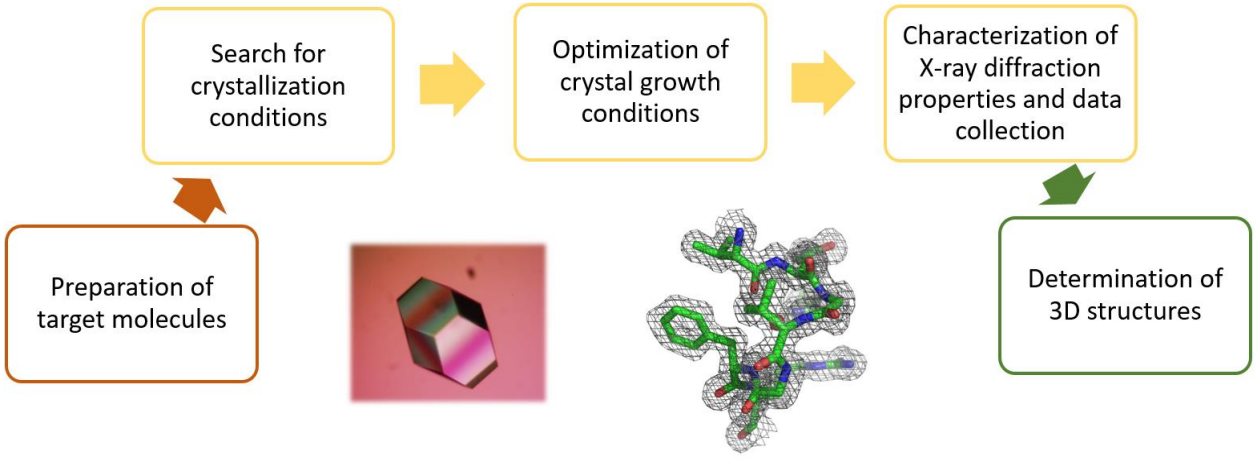


Figure 1. The schematic protocol employed in a biocrystallographic project.

II. Protein-ligand interactions

Proteins play a vital role in living systems and are very complex molecules. Typically, the amino-acid sequence shapes the protein structure and function but very often to perform their work a protein needs a "little help" from different molecules. The 3D structure of such proteins usually shows a distinct binding site for these molecules. Any molecule bound to a protein and forming a complex with it is referred as a ligand (from Latin *ligare* meaning *to bind*). Ligands include substrates, inhibitors, activators and neurotransmitters and they can be peptides, proteins, DNA, small molecules, ions, and drugs.

A well-known example would be hemoglobin, which is a protein responsible for the oxygen transport in vertebrates. This would not be possible without iron atoms. Each hemoglobin carries four distinct, non-covalently bonded heme groups – each one accommodating one iron atom. Once it binds to dissolved oxygen atoms, the protein can transport oxygen from lungs to tissues - graphical representation shown in Figure 2.

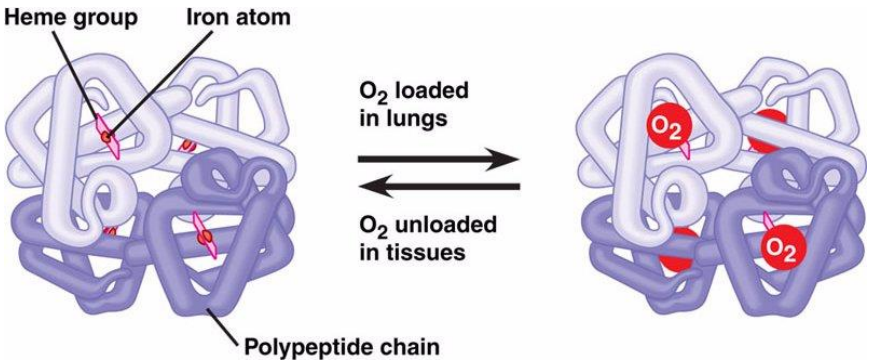


Figure 2. Hemoglobin mode of action. Each hemoglobin contains four distinct iron-containing heme groups. Oxygen is loaded during respiration. Once the tissue is reached deoxygenation takes place, and metalloprotein can continue the cycle. Image taken from stepwards.com website (access on May 8th 2017).

There are plenty of such examples, and each would have a unique mechanism. The order of interaction between the protein and a ligand lays within weak, non-covalent bonds such as salt bridges, hydrogen bonds, and *Van der Waals* forces. The binding site specificity is related to the nature of R-groups (amino acids side chains) mutual interaction, giving rise to a binding pocket within the protein structure.

There are 20 common amino acids, very different in chemical character, thus being able to form different bonds. This is why those interactions are usually very specific.

There are many ways for amino-acids classification, and grouping them by their chemical character is the most convenient way to visualize types of a bond they can take part in. Hydrophobic interactions and *Van der Waals* forces stand for an important class in protein science, thus giving a significant role to non-polar residues, such as aliphatic (Gly, Ala, Val, Leu, Ile, and Pro) and aromatic (Phe, Tyr, Trp) ones. Non-polar groups contribute to change in entropy, since they are more stable when close to each other rather than exposed to water – this would be a basic description of hydrophobic interactions. *Van der Waals* forces are related to the charge of a molecule or more accurately - charge fluctuation. This behavior results in temporary charge attraction or repulsion of the neighboring electrons and atoms - eventually giving rise to the attractive forces.

Charge-associated interactions would include acidic (Asp, Glu) and basic (Arg, Lys, His) residues (here charge-charge and ionic interactions). Ionic interactions tend to be stronger when considering amino acids containing alcohols (Ser, Thr, Tyr) or amides (Asn, Gln) because of the presence of either oxygen or nitrogen atom and their electron withdrawing properties. Their polarity makes them a good fit for hydrogen bonding, between N-H, O-H hydrogens, and a lone pair on oxygen and nitrogen at an adjacent residue.

Another important class of bonds in proteins are those forming disulfide bridges between Cys residues. Those are formed either within a peptide or between separate polypeptides. They form a covalent bond, meaning they need a catalyst to be formed or broken apart. They are responsible for many catalytic properties of certain enzymes, and an example is also given in this work.

III. Aim of the study

This thesis is based on the structural characterization of the protein-small molecule interactions in two independent systems:

1. Glutathione reductase (GR) from the photosynthetic algae *Chlamydomonas reinhardtii*, which binds a flavin adenine dinucleotide (FAD) in each chain. This subject will be referred as PROJECT1 in this thesis.
2. Bovine serum albumin (BSA) as a putative system to solubilize and carry in water solutions boron-dipyrromethene (BODIPY), a novel fluorescent dye. This subject will be referred in this thesis as PROJECT2.

PROJECT1: Glutathione reductase plays a significant role in the cell homeostasis by maintaining a reducing environment within a cell. In photosynthetic organisms it is present in several cell compartments, thus giving rise to different isoforms. Even though its function and mode of action are believed to be well conserved in nature, there is little known about this structure of plant origin.

In Protein Data Bank (PDB) at the moment there are 36 structures of GR, but any of these from photosynthetic organisms.

The target of this project is to determine the three-dimensional structure of two GR isoforms, namely glycolytic (GR1) and chloroplastic (GR2), characterize the interaction of the protein with FAD with a focus on Rossman fold holding the cofactor and to perform a comparative study with other GRs of different origin already characterized.

This work is a collaborative project between plant physiology and biocrystallography group at the University of Bologna. The first one being responsible for biochemical characterization (including expression and purification) and latter for obtaining high-resolution structure and probing interactions. Once purified to homogeneity, the protein is taken for crystallization trials. If positive results are obtained, crystallization conditions have to be optimized to get a representative, high-quality crystals. Those are then taken to macromolecular crystallography beamline at Synchrotron light source for the diffraction data collection.

PROJECT2: Fluorophores having excellent photophysical characteristics were always in demand in every area of natural and life sciences. First syntheses and characterization of boron-dipyrromethene fluorescent dyes were performed in 1968 but gained popularity quite recently. Their characteristics were real *El Dorado* in fluorescent tools as stated by Ziessel et al. (2007).

Due attractive fit in cell imaging it became an interesting aspect to investigate BODIPY interaction with proteins. One of the drawbacks of these molecules is their very low solubility in water, and their use in complex with proteins could be a smart way to overcome this problem.

Thus, the aim of this project is to develop a procedure for preparation of protein-BODIPY adduct and characterize it by a spectroscopic and crystallographic approach for probing protein-small molecule interactions. The choice of the protein was addressed toward proteins commercially available and well characterized as Bovine serum albumin (BSA).

Emphasis is on spectroscopic characterization to investigate photophysical properties of the dye when incorporated in protein (optical spectroscopy approach) and structural insights to see the range of interactions and explore potential binding pocket (biocrystallographic approach). Applying hybrid methods gives a complete image of protein-small molecule behavior regarding photophysical and chemical stability, being a good starting point for further research in the area.

PROJECT1

Chapter 2 - Introduction

2.1 Biological importance and catalytic mechanism of Glutathione Reductase

Abundant in most organisms Glutathione Reductase (GR) or Glutathione Disulfide Reductase (GSR) is an enzyme that is conserved among all kingdoms, it has a well-defined role in cells and for a number of organisms, its enzymatic activity mechanism is well understood. Its key role is to resist an oxidative stress when needed and to control reducing environment of the cell. Indeed, it catalyzes the reduction reaction of glutathione disulfide (GSSG) to glutathione (GSH). [1]

Glutathione (γ-L-Glutamyl-L-cysteinyl-glycine) is a water soluble tripeptide. GSSG consists of two GSH molecules linked by the disulfide bridge, as shown in Figure 3A. Glutathione as a ubiquitous thiol-containing tripeptide participates in its reduced form (GSH) in several functions of vital importance to the cell. [2]

GR is a homodimer employing flavin adenine dinucleotide (FAD) and nicotinamide adenine dinucleotide phosphate (NADPH) to efficiently convert one molar equivalent of GSSG into two equivalents of GSH. The schematic mechanism is based on proton donation/acceptance as in Figure 3B. The enzymatic mechanism of GR is of ping-pong type and is connected to the reductive and oxidative half reaction. This process is facilitated by a few conserved residues which are discussed in details in proceeding sub-chapter.

GR catalytic activity is NADPH-dependent process involving cysteine (Cys) residues. Disulfide reduction reaction of GSSG to GSH is initiated by FAD through the nucleophilic attack. The process is based on sulfur biochemistry, thus it requires an electron relay that would link reducing NADPH with thiol/disulfide and thiol containing adapter molecule to transfer an electron to different acceptors (here GSH). [3]

2.2 Glutathione Reductase role in eukaryotes

The structure and function of GR in pro- and eukaryotes is well known for years. GR in plant eukaryotes is found in chloroplasts, mitochondria, and cytosol.

From functional aspect GR catalytic properties are essential for efficient removal of dangerous Reactive Oxygen Species (ROS) being, for instance, hydroxyl radicals ($\cdot OH$), singlet oxygen (O_2^1) and other electrophiles. In human cells it was reported that GR function is in fact much more than that, being responsible for metabolism and clearance of xenobiotics, regenerating antioxidants (such as vitamin C and E) to their active forms and acting as a cofactor in detoxifying enzymes. [1]

GR was first purified from baker's yeast in 1955 by E. Racker at Yale, who also described electron donor properties of NADPH in this enzyme. Ten years later his group also verified a presence FAD and thiol group leading to the first mechanism proposal. Since then many molecular insights of this enzyme had been unrevealed along with structure determination at 3 Å resolution by Schulze *et al.* (1978) and up to now it remains an extensively studied flavoprotein. [4]

2.3 Abiotic stress defense in plants

ROS are generated in living organisms as byproducts of causal metabolic pathways. Abiotic and biotic stress is considered to be of critical importance in the maintenance of a cell cycle and homeostasis. Abiotic stress by simplest definition is a negative influence of non-living factors on the organism at specific conditions (here environment). This includes drought, flooding, radiation, temperature, nutrient access and wind. Each of these is unavoidable to some extent when we consider plants, and they speed up ROS formation in a cell. At low ROS levels, the cell performs its vital functions, such as cell growth, pathogenic defense and/or development. When this is not the case - thus ROS and oxidants are exceeding the levels of vital antioxidants - the resulting events may be severe for a cell, leading to protein, lipids, carbohydrates and/or nucleic acids damage. Not surprisingly this may cause cell malfunction and death.

Plants have developed powerful metabolic pathway in preventing ROS-related reactions – ascorbate-glutathione (AsA-GSH) cycle. The process starts with hydrogen peroxide

being reduced to water, utilizing ascorbate peroxidase (APX) with an ascorbate as an electron donor. Oxidized ascorbate radical gets reduced by monodehydroascorbate (MDAR) reductase but since it is a radical the reduction has to be very efficient, otherwise, dehydroascorbate and ascorbate are formed. This is the point where GSH is crucial to effectively convert dehydroascorbate into ascorbate with help of dehydroascorbate reductase (DHAR). After reduction step, GSH is oxidized to GSSG. Furtherly GSH is recovered by glutathione reductase, requiring NADPH as an electron donor.

The GSH/GSSG ratio has to be normally maintained in favor of GSH to effectively get through abiotic stress.

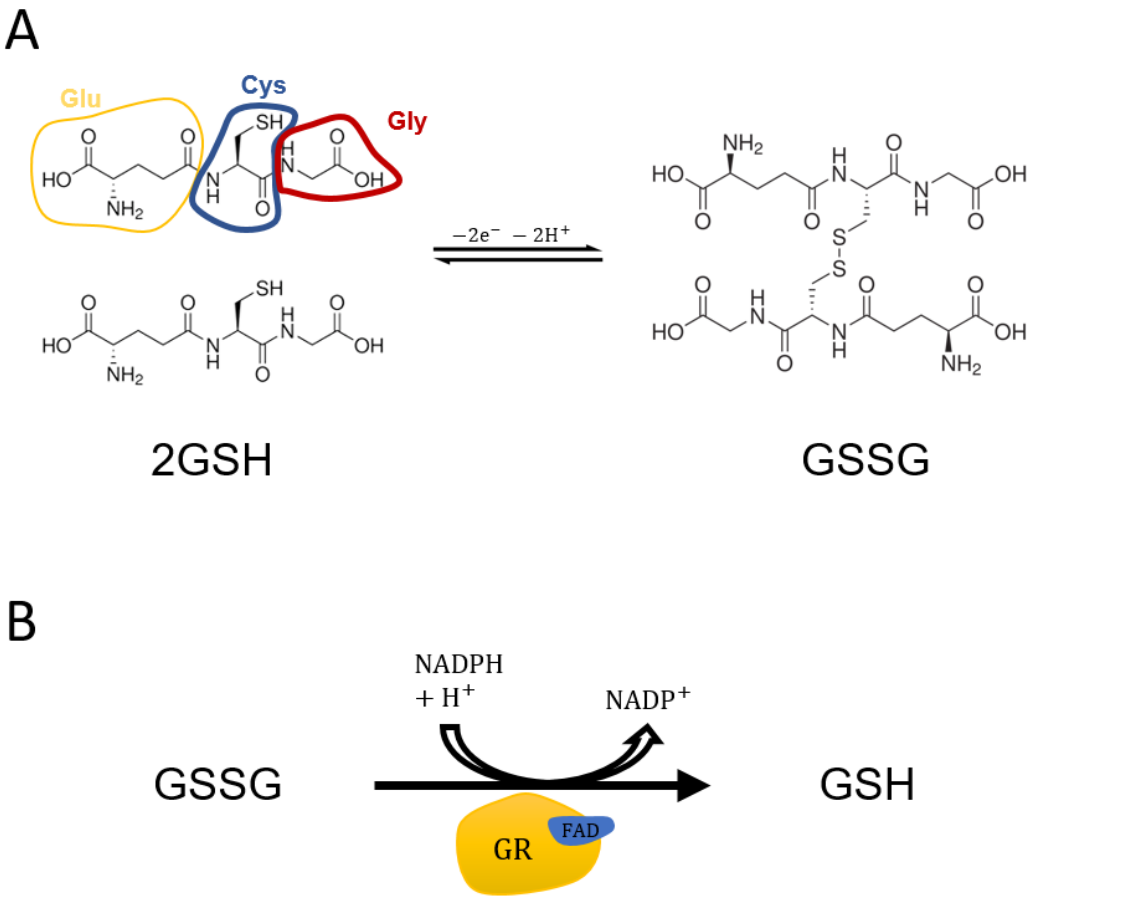


Figure 3. Scheme of redox reaction upon GR catalytic activity. (A) Simplified representation of sulfur biochemistry in 2GSH/GSSG. Withdrawal of electrons from thiol groups leads to formation of sulfur bridge, thus oxidized form of glutathione. (B) Global representation of redox reactions with participation of FAD-bound GR and accompanying oxidation of NADPH.

Flavoproteins plays a role of an electron relay, and reducing NADPH hence enters the cycle with the guidance of GR-FAD complex yielding 2 molecules of GSH. The scheme of this cycle is shown in Figure 3B. GSH furtherly serves as a reducing agent for disulfides and hydroperoxides or conjugates with 2-oxoaldehydes (2-OA) and electrophiles.
Summing up, disulfide reducing properties of GR makes it of critical importance in removal and/or conversion of electrophilic substances and xenobiotics. This process also demonstrates the adaptability of glutathione-based catalysis as an explanation of various chemical challenges in our understanding of evolution. [1]

2.4 Glutathione reductase structure and enzymatic characteristics

As shortly mentioned before GR crystal structure was determined in 1978 at 3 Å resolution by Schultz *et al.* and its characteristics was advancing through sequence assignment in 1981 (from crystallographic data at 2 Å by Thieme *et al.*), genetic screens and redox catalysis. Today GR is one of the best-understood enzymes. [5]
GR is a flavoenzyme belonging to the family of pyridine-nucleotide-disulfide-oxidoreductases. Despite the fact that it is found in certain isoforms in both pro- and eukaryotes it forms a stable homodimer with an approximate mass of ~110 kDa shown in Figure 4A. Each subunit contains two distinct Rossman folds accommodating FAD and NADPH. The flavin ring of a FAD molecule separates the binding sites for NADPH and their spatial orientation within an enzyme is presented in Figure 4B. In a case of GSSG, the binding takes place at the interface between GR monomers, implying that this enzyme is functional only in its homodimeric form. The cofactor binding domains are known from GR crystal structures of *E. Coli*, *H. Sapiens*, and *S. Cerevisiae*, where interface domain is known to be responsible for dimerization and comprises two regions involving residues located at C- and N-terminus of the protein. Significant sequence variation in GR structures occurs around position 75 (interface domain). In this region, *E.Coli* GR has an antiparallel intersubunit β-sheet from 71 to 80, *H. Sapiens* GR has an intersubunit disulfide bridge (C90–C90'), and *S. Cerevisiae* GR sequence from 73 to 88 formed more tight extended intersubunit antiparallel α-helices. The most important feature of the crystallized GRs is that they all contain an intersubunit in the active site. This bridge is opened when

NADPH binds to the oxidized enzyme. In this form, GR is capable of attacking GSSG and a mixed disulfide is formed.

It is commonly believed that GR structure by means of both substrate binding and the amino-acid sequence is well conserved in the course of evolution among all kingdoms. A small distinction is assumed for the subunit interface, yet as demonstrated in proceeding chapters plant GR enzyme differs quite significantly from human isoforms. Least conserved residues for a variety of organisms are Cys residues, playing usually regulatory role. [6]

In the context of function, GR being an oxidoreductase has three substrates, namely NADPH, H^+ and GSSG and two products (2 molecules of GSH). Schematically its role in metabolism was already presented in Figure 3B. A similar observation can be withdrawn when considering isoforms balance, which for instance in yeast was reported to be regulated by translation initiation efficiency. [7]

GR high resolution structure has been reported for certain pro- and eukaryotic organisms varying from 1.55 Å up to 2.6 Å. It is essential to know sequence and structure similarities in order to confirm interactional features within an enzyme. Based on PBD search 8 structures from different organisms were taken under consideration being namely: *Yersinia pestis* (gram-negative bacteria), *Streptococcus mutans* (gram-positive bacteria), *Vibrio parahaemolyticus* (gram-negative bacterium), *Bartonella henselae* (proteobacterium), *Saccharomyces cerevisiae* (yeast), *Plasmodium falciparum* (protozoan), *Escherichia coli* (gram-negative bacteria) and *Homo sapiens*. This simply means there is no single structure being reported for plants.

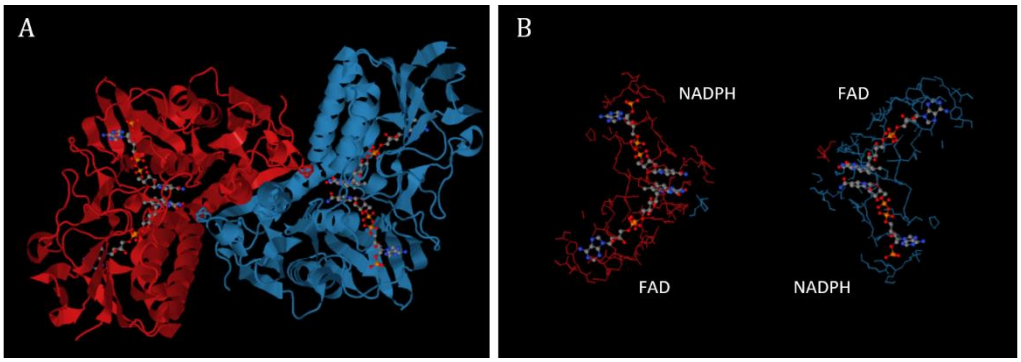


Figure 4. GR homodimer with its binding sites from *Escherichia coli*. (A) Subunit representation of an enzyme and (B) FAD and NADPH spatial orientation within GR with surrounding residues. The catalytic activity is located at the interface between GRs monomers. Images generated with PDB viewer, PDB ID: 1GET

2.4.1 Reductive half-reaction of Glutathione reductase

Oxidized state GR (GR_{ox}) involves two cysteine (Cys47B and Cys42B for bacterial GR) residues forming disulfide bridge on the 'back' side of isoalloxazine ring of the FAD. The noteworthy disulfide bridge is in the proximity of histidine residue (His439A) which is hydrogen bonded to glutamic acid (Glu445A) – where both of these residues belong to the A subunit of GR homodimer. On the other side (facing NADPH) of flavin, there is also tyrosine residue which acts as a shield for FAD and stands for a gate-keeper at NADPH site. Once NADPH site is filled the tyrosine ring rotates away from flavin ring and locks the nicotinamide group of NADPH. Then hydride from NADPH reduces FAD to $FADH^-$ which furtherly transfer electron pair to the neighboring cysteine residue (Cys47B). This gives rise to stable charge transfer complex with a flavin ring. The other Cys residues are protonated by a histidine residue. [8][9] Up to this point, there is two-electron reduced enzyme (GRH_2), NADP+ is then dissociated leaving space for another NADPH molecule. [10] Scheme of a simplified mechanism of this redox cycle is shown in Figure 5.

2.4.2 Oxidative half-reaction of Glutathione reductase

Oxidative part of enzymatic catalysis of GR starts with GSSG binding to reduced enzyme GRH_2 accompanied by tyrosine residue reorientation, so that its hydroxyl group is pointing at disulfide bridge of GSSG. It also stabilized by other conserved residues from both subunits of homodimer – four positively and two negatively charged residues compensating substrate charge. Then the substrate is attacked by interchange residue Cys_{int} (Cys47B) which in consequence forms a mixed disulfide bond. The His (His439A) residue withdraw the proton from Cys_{int} accelerating the nucleophilic attack overall. Protonated His will then give off its proton in favor of one thiolate group of GSSG that is being reduced. The reaction is supposed to take place in accompany of Tyr residue. [9][11] Once the GSH group detaches from the active site, the mixed disulfide bond is being attacked by sulfur atom of Cys_{CTC} (CTC - Charge Transfer Complex) oxidizing GR back again. Second GSH molecule gets a proton from same His residue closing the cycle. The rate of these reactions was reported to be not constant in time, since one of the

protonation steps serve as a rate limiting step in terms of kinetics. [11][12] Nonetheless His plays influential role in GR enzymatic catalysis.

Current understanding of this process is rather well established, but still, there are many open questions about mechanistic details, including source and fate of protons, role of Tyr residue, which Cys residues are protonated by His and what is the order to this redox process, is it sequential or does it happen simultaneously?

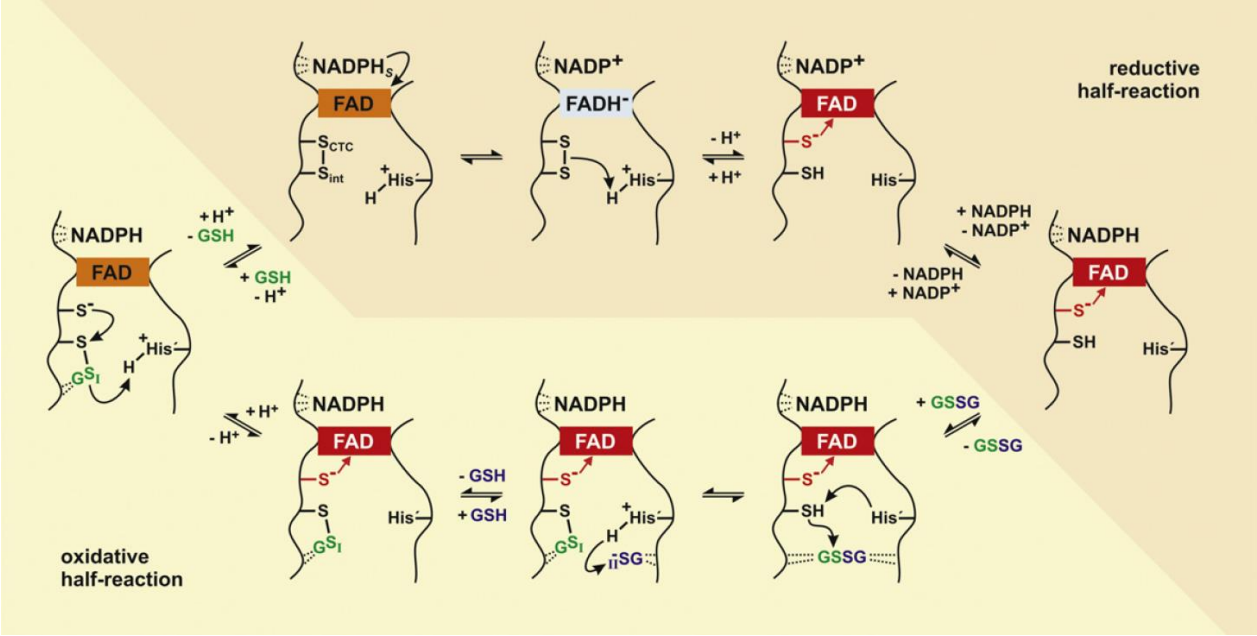


Figure 5. Schematic representation of oxidative and reductive half reaction of GR. Image taken from [1].

2.5 *Chlamydomonas reinhardtii* as a model organism for GR studies

C. reinhardtii is a single-cell, eukaryotic green alga being soil and freshwater abundant organism. Its cell wall is made of hydroxyproline-rich glycoproteins and has a large cup-shaped chloroplast with light-sensible eyespot. It is a very popular model organism in biology due to ease of culturing and simplicity of genetic manipulation. It also has notable cell cycle feature of growing photoautotrophically under light illumination and in the dark if organic carbon is supplied. Many questions in molecular biology were resolved with *C. reinhardtii*, demonstrating how do cells move, behave under light irradiation, recognize one another or react to changes in mineral nutrition. From the biotechnological view, it is a meaningful organism in biopharmaceuticals and biofuel production as well as a tool in hydrogen generation. [13]

There are many mutants of *C. reinhardtii* and they are extensively used for a study of various biological processes, like flagellar motility, photosynthesis or protein synthesis. These species are normally haploid so that the effects of mutations are seen immediately without further crosses. Those facts make this organism a very good fit for a comparative study of plant-based enzymes.

In *C. reinhardtii* GR is present in all cellular compartments, such as chloroplasts, mitochondria, endoplasmic reticulum, vacuoles, and cytosol. Chloroplast and mitochondrial GR is encoded by gr2, while gr1 encodes a protein that is found in the cytosol and peroxisomes (Kataya and Reumann, 2010). In *Chlamydomonas reinhardtii* there are two genes encoding GR: the product of GSHR1 is cytosolic, and the product of GSHR2 is plastidial (Merchant *et al.*, 2007).

Chapter 3 - Methods and experimental fine points

This chapter deals with protein preparation, protocols and basic instrumental insights needed to determine the protein structure. Protein preparation, thus expression and purification were performed by the Plant Physiology group at Department of Pharmacy and Biotechnology (FaBit) of the University of Bologna.

3.1 Protein preparation and purification

The pET3 plasmids containing the cDNAs encoding *Chlamydomonas reinhardtii* proteins glutathione reductase 1 (GR1) and glutathione reductase 2 (GR2) were kindly provided by Dr. Stephane D. Lemaire from the Institut de Biologie Physico-Chimique (IBPC) in Paris (France).

Plasmids were introduced into BL21 *E. Coli* competent cells and positive clones from antibiotic selection were cultured in the LB (Luria-Bertani) medium. Culturing was done at 37 °C, over 16-18h with continuous shaking. About 8 mL of culture was transferred into 800 mL of fresh LB medium and again grown at 37 °C until an optical density of 0.4-0.5 at 600 nm was reached. Protein expression was induced by 0.2 mM IPTG and left for 3 hours for GR1 and for 16-18 hours for GR2 (at 30 °C). Then cells were harvested by centrifugation (8000 xg for 20 minutes), the pellet was washed with buffer and pelleted again before storage at -80 °C. Frozen pellet was re-suspended in binding buffer (20 mM Tris-HCl, 0.5 mM NaCl, 5 mM imidazole, pH 7.9) and taken for 3 cycles of French Press. Cell debris was removed by centrifugation and supernatant containing recombinant enzyme was recovered. Purification was done with Ni-NTA affinity column and after extensive washing with step-gradient concentration of imidazole (30 and 60 mM), the elution was performed in the presence of 250 mM imidazole. Eluted fractions were desalted using PD-10 columns (GE-Healthcare) equilibrated with Tris-HCl 30 mM, pH 7.9. Purified recombinant proteins were concentrated using Centricon filters and stored at -20 °C.

The electrophoretic analysis was done with SDS-PAGE with 12% acrylamide gels and stained with Coomassie Blue (40% ethanol, 12% acetic acid, 0.25% Coomassie Blue R-250). Bleaching was done with destaining solution (40% ethanol, 10% acetic acid).

3.2 Dynamic Light Scattering

To check the oligomeric state of the protein, Dynamic Light Scattering (DLS) measurements were performed. DLS is a technique that measures the dimension of particles in solution, in a submicron range. The measurement is based on relating Brownian motion to the size of a particle. The larger the particle the slower the Brownian motion. For the DLS measurement, the temperature has to be accurately known since it requires viscosity information for data evaluation. The velocity of the Brownian motion is defined by the translational diffusion coefficient. For this reason, it is essential to keep the temperature stable throughout the experiment.

The DLS spectrophotometer is constituted by a laser light source illuminating the sample contained in the cell, which causes the scattering of light recorded by the detector. It measures the time-dependent fluctuations in scattered light from a solution of molecules to determine the hydrodynamic radius (R_H) from the Stokes-Einstein equation

$$\mathrm{d(H)} = \frac{\mathrm{kT}}{3\pi\eta\mathrm{D}}$$

Where:

$\mathrm{d(H)}$ – hydrodynamic diameter

D – translational diffusion coefficient

k – Boltzmann's constant

T – absolute temperature

η – viscosity

The diameter in DLS measurement is referred as how particle diffuse in a fluid thus it is named hydrodynamic diameter. This statement directly entails other parameters to be of high importance such as concentration and chemical character of a sample. The translational diffusion coefficient is dependent in fact on the size of the particle, ionic strength of the medium (affecting diffusion speed), surface structure (surface change can affect apparent size of the particle), and non-spherically shaped particles.

DLS from a physical point of view is dominated by the scattering phenomenon. This can be split into two elementary types of scattering being namely Rayleigh and Mie scattering. The first one says that if the particles are small compared to the wavelength of the laser used (typically ~600nm), then the light scattered from the particle will be isotropic. The

Rayleigh theory tells us that intensity of light scattered is inversely proportional to the particle diameter to the power of six

$$I \propto d^6$$

and

$$I \propto \frac{1}{\lambda^4}$$

In practice, this means that if there is a hypothetic particle 50 nm in size it will scatter 10^6 as much light as a 5 nm particle. This implies a conclusion that light from large particles will suppress the light scattered from the small ones. Mie concept of scattering says that if the size of particles become similar to that of the wavelength of incoming light then a complex function of maxima and minima in respect to the angle is observed.

The DLS measurement to be reliable needs to be correlated with either two independent signals or same signal in a time domain in search for similarity degree. If the intensity of a signal is compared with itself at a particular point in time and a time much later, then for a randomly fluctuating signal it is obvious that the intensities are not going to be related in any way. Knowledge of the initial signal intensity will not allow the signal intensity at time $t = \infty$ to be predicted. However, if the intensity of a signal at time = t is compared to the intensity a very small time later ($t + \delta t$), there will be a strong relationship or correlation between the intensities of two signals. The two signals are strongly or well correlated. If the particles are large the signal will be changing slowly and the correlation will persist for a long time. If the particles are small and moving rapidly then the correlation will reduce more quickly. [14]

The DLS measurements were done with a Zetasizer Nano ZS, Malvern, with a laser source He-Ne (633 nm, 5 mW) at a temperature of 25 °C, in a 100 µL cuvette with 30 mM Tris-HCl 30 mM (pH 7.9), and the concentration of analyzed samples was 1 mg/mL for each protein (CrGR1and CrGR2). The correlator compares the intensity of scattering in subsequent time intervals and sends the data to a software to be analyzed. The reported hydrodynamic radii (R_H) for each protein sample have been averaged from the values obtained from five measurements each one composed of ten runs of 10 seconds.

3.3 Crystallization

In XRD experiment it is crucial to have high-quality crystals. To achieve that many strategies have been explored with varying outcomes. Normally the protein purity is essential along with its properties under physiological conditions (such as solubility and activity). These properties make often crystallization rather unpredictable, thus requiring trials, and optimizations if satisfactory conditions are found.

Crystallization is a process purely dominated by the laws of the physical chemistry. In this context, there is no difference in crystallizing sodium chloride (NaCl) or macromolecular complex. The difference relies on a method of bringing the molecule of interest to the supersaturation level. While for low molecular weight compounds (either organic or inorganic) this is not considered a problem, for proteins the supersaturation level cannot be reached for example by simple solvent evaporation or extensive heating and dissolving. To crystallize means to gradually decrease the solubility of an analyte so that it will precipitate and eventually transition from liquid to solid phase would result in a crystalline material. Schematic phase diagram for protein is presented in Figure 6. The undersaturated region stands for a region of the soluble protein. Then the solution overcome the solubility curve entering the metastable region and even though in the supersaturated conditions the nucleation cannot take place, but when the crystalline nucleus are formed they can grow crystals. The labile zone is a region when crystals can nucleate and growth takes place quicker and is less stable. The last step in crystallization apart from nucleation and crystal growth is termination step. It takes place when supersaturation is lost due to solid phase formation and solubility curve is reached.

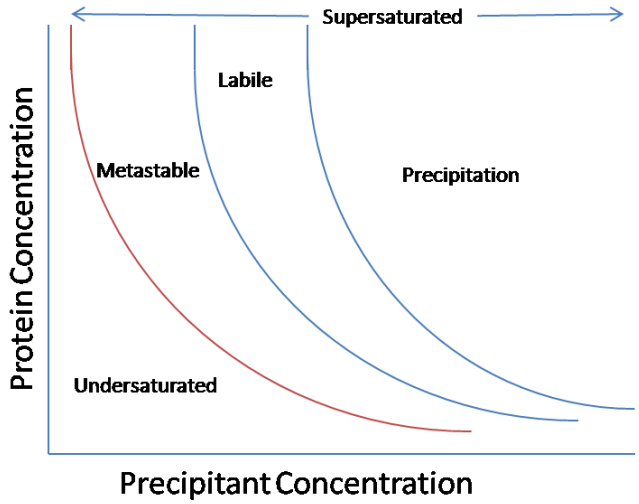


Figure 6. Phase diagram for generic protein.

Crystallization of proteins is normally facilitated through addition of several compounds affecting the physicochemical character of the solution, like for instance protein concentration, precipitating agent nature and concentration, ionic strength, pH,

temperature and dielectric constant of the solvent. The precipitant promotes macromolecule aggregation, by reducing solvation shell of the protein since they will compete for water molecules. To reach this there are few groups of compounds utilized, such as salts, polymers, and organic solvents. For example, salts will increase the ionic strength due to dissociation and will stabilize protein through interaction with solvent-accessible charged residues. Polymers increase molecular crowding. Polymer and protein will compete for water molecules brought by hydrogen bonds with hydroxyl moiety and eventually amplify protein-protein interaction. A similar mechanism is accepted for organics that withdraw the water molecules form solvation shell of the macromolecule. Next important compositional feature of crystallization solution is the buffer that determines pH. Its value affects the charge distribution on a protein surface (pK_a and pI values). The pH range is also critical for protein stability. The crystallization solution composed of precipitant, buffer and salt/additive is commonly referred as a reservoir solution. Last crucial parameter in crystallization is the temperature since it directly affects the protein solubility and crystal growth rate. It is necessary to keep it constant to have controlled crystallization growth conditions.

Up to now, there are certain methods for carrying out crystallization step and this includes vapor diffusion, micro batch, dialysis and free-interface diffusion.

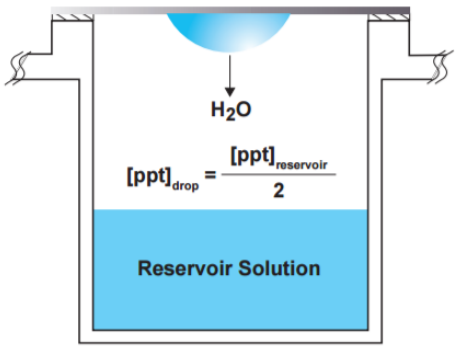


Figure 7. Scheme of a hanging drop vapor diffusion method.

Vapor diffusion is found to be very effective method for macromolecule crystallization and rely on a straightforward principle. The protein sample and respective reservoir solution chosen for crystallization are placed on a siliconized coverslip (in a small volume) and left for vapor equilibration with ‘mother solution’ being reagent mixture in sizeable volume (with respect to the drop). The scheme of such setup is presented in Figure 7 for the case of a hanging drop. The equilibration is finalized when the concentration of the reagent in the drop is equal to the one in the reservoir (here water leaving the drop). As the concentration of protein gradually increases it eventually reach supersaturation domain and starts to either

crystallize or precipitate – salting out. This strategy is cost effective, allows for many experiments at once and gives easy access to the crystals. [15]

Batch crystallization is based on mixing the protein sample with reservoir solution in a small amount to form a micro-drop that is then covered with an oil to avoid or limit evaporation. The oil diffusive properties are varying, for example covering micro-drop with paraffin oil reduce the diffusion completely while AI's oil (1:1 mixture of silicon and paraffin oil) will permit some extent of diffusion, hence bringing protein to the supersaturated state.

In dialysis crystallization method, the protein solution is equilibrated against a large volume of reservoir solution through a membrane. The gradual increase of precipitant concentration will eventually lead to nucleation zone and crystal growth of the protein.

Last, most effective (probably) and not accessible method is a crystal growth in microgravity conditions. The advantage of this approach is the reduction of a convective motions and sedimentation, meaning the crystal growth is more uniform, fewer contaminants present within the crystal and most importantly the quality of the crystals was reported to be of outstanding quality. Many successful and interesting examples of this method are listed in [16].

Crystallization experiments were manually performed by the hanging drop vapor diffusion method, using a 24-well Linbro plate from Molecular Dimension. For the trial crystallization Structure Screen 1 cacodylate free from Molecular Dimensions had been used (MD1-01-CF). It contains 50 screening solutions and most of them were tested. The reservoir solution compositions of MD1-01-CF is presented in Table 1 and extensive kits can be found in an additional information 1. For optimization, two extra kits were applied: JCSG-plus (MD1-37) from Molecular Dimensions and Crystal Screen 2 from Hampton Research by picking a tailored composition. Some optimizations required manual preparation and are indicated as 'manual' in this work.

The protein solution was maintained at a concentration of 10 mg/mL for all crystallization trials and was suspended in buffer and 2% EDTA solution. 2 µL of protein was placed on a siliconized coverslip and mixed with another 2 µL of mother solution from the reservoir (700 µL). Each well was vaseline-sealed and left in an incubator at 20 ± 1 °C. Control check was done every week under an optical microscope (Leitz Wetzlab SM-LUX) equipped with a digital camera Motic 5.0 MP. Figure 8A presents prepared plates for GR1

and 8B for GR2 for crystal screen and Figure 8C presents optimized conditions for GR1. A number of crystallization trials differ for GR1 and GR2 due to different volumes available at given time.

Table 1. Structure Screen 1 cacodylate free - tube composition used for crystallization trials.

Tube#	Salt	Buffer	pH	Precipitant
1	0.02 M calcium chloride	0.1 M sodium acetate	4.6	30% v/v MPD
2	0.2 M ammonium acetate	0.1 M sodium acetate	4.6	30% w/v PEG 4K
3	0.2 M ammonium sulfate	0.1 M sodium acetate	4.6	25% w/v PEG 4K
4	None	0.1 M sodium acetate	4.6	2.0 M sodium formate
5	None	0.1 M sodium acetate	4.6	2.0 M ammonium sulfate
6	None	0.1 M sodium acetate	4.6	8% w/v PEG 4K
7	0.2 M ammonium acetate	0.1 M tri-sodium citrate	5.6	30% w/v PEG 4K
8	0.2 M ammonium acetate	0.1 M tri-sodium citrate	5.6	30% v/v MPD
9	None	0.1 M tri-sodium citrate	5.6	20% v/v 2-propanol, 20% w/v PEG 4K
10	None	0.1 M tri-sodium citrate	5.6	1.0 M ammonium dihydrogen phosphate
11	0.2 M calcium chloride	0.1 M sodium acetate	4.6	20% v/v 2-propanol
12	None	0.1 M MES	6.5	1.4 M sodium acetate
13	0.2 M tri-sodium citrate	0.1 M MES	6.5	30% v/v 2-propanol
14	0.2 M ammonium sulfate	0.1 M MES	6.5	30% w/v PEG 8K
15	0.2 M magnesium acetate	0.1 M MES	6.5	20% w/v PEG 8K
16	0.2 M magnesium acetate	0.1 M MES	6.5	30% v/v MPD
17	None	0.1 M imidazole	6.5	1.0 M sodium acetate
18	0.2 M sodium acetate	0.1 M MES	6.5	30% w/v PEG 8K
19	0.2 M zinc acetate	0.1 M MES	6.5	18% w/v PEG 8K
20	0.2 M calcium acetate	0.1 M MES	6.5	18% w/v PEG 8K
21	0.2 M tri-sodium citrate	0.1 M HEPES	7.5	30% v/v MPD
22	0.2 M magnesium chloride	0.1 M HEPES	7.5	30% v/v 2-propanol
23	0.2 M calcium chloride	0.1 M HEPES	7.5	28% w/v PEG 400
24	0.2 M magnesium chloride	0.1 M HEPES	7.5	30% w/v PEG 400
25	0.2 M tri-sodium citrate	0.1 M HEPES	7.5	20% v/v 2-propanol
26	None	0.1 M HEPES	7.5	0.8 M K/Na tartrate
27	None	0.1 M HEPES	7.5	1.5 M lithium sulfate
28	None	0.1 M HEPES	7.5	0.8 M sodium dihydrogen phosphate, 0.8 M K dihydrogen phosphate
29	None	0.1 M HEPES	7.5	1.4 M tri-sodium citrate
30	None	0.1 M HEPES	7.5	2% v/v PEG 400, 2.0 M ammonium sulfate
31	None	0.1 M HEPES	7.5	10% v/v 2-propanol, 20% w/v PEG 4K
32	None	0.1 M Tris	8.5	2.0 M ammonium sulfate
33	0.2 M magnesium chloride	0.1 M Tris	8.5	30% w/v PEG 4K
34	0.2 M tri-sodium citrate	0.1 M Tris	8.5	30% w/v PEG 400
35	0.2 M lithium sulfate	0.1 M Tris	8.5	30% w/v PEG 4K
36	0.2 M ammonium acetate	0.1 M Tris	8.5	30% v/v 2-propanol
37	0.2 M sodium acetate	0.1 M Tris	8.5	30% w/v PEG 4K
38	None	0.1 M Tris	8.5	8% w/v PEG 8K
39	None	0.1 M Tris	8.5	2.0 M ammonium dihydrogen phosphate

40	None	None	-	0.4 M K/Na tartrate
41	None	None	-	0.4 M ammonium dihydrogen phosphate
42	0.2 M ammonium sulfate	None	-	30% w/v PEG 8K
43	0.2 M ammonium sulfate	None	-	30% w/v PEG 4K
44	None	None	-	2.0 M ammonium sulfate
45	None	None	-	4.0 M sodium formate
46	0.05 M potassium dihydrogen phosphate	None	-	20% w/v PEG 8K
47	None	None	-	30% w/v PEG 1.5K
48	None	None	-	0.2 M magnesium formate
49	0.2 M lithium sulfate	None	-	2% w/v PEG 8K
50	0.2 M lithium sulfate	None	-	15% w/v PEG 8K

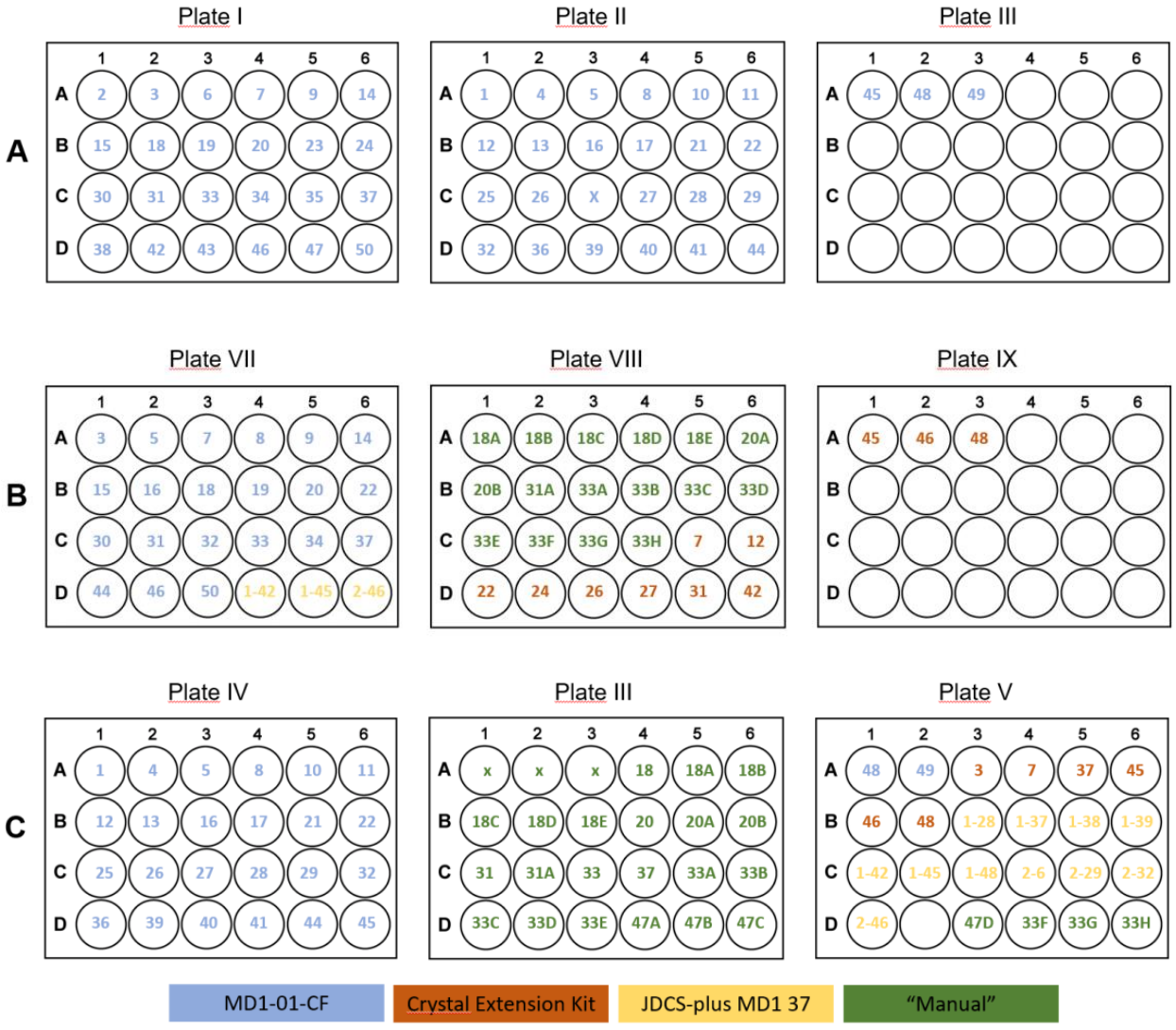


Figure 8. Scheme of well numbering with a color indication respective to the crystallization kit used for GR1 and GR2 crystallization trials and optimizations.

3.4 Principles of Single Crystal X-ray Diffraction

3.4.1 X-ray Diffraction

Single crystal X-ray diffraction is an analytical technique used for determining the atomic and molecular structure of a crystal. Crystalline material causes diffraction of an X-ray beam directed onto it. Once the crystal is grown it is mounted on a goniometer in order to place it in certain positions with respect to the incident beam (exemplary image provided in discussion chapter). Gradual rotation of a crystal gives a set of specific diffraction patterns. When focused, monochromatic and, coherent X-ray beam hits the crystal it will scatter light in many directions, appearing on a detector as regularly spaced spots – called reflections. Each of those is located at a specific position from the origin and has some intensity. Spots are named reflections because each of them corresponds to an X-ray beam reflected from an individual and evenly spaced planes within a crystal. Since the image is two-dimensional, and the ultimate aim is to get three-dimensional electron density map of the molecule determining the scattering, those images are taken at many different crystal orientations.

The X-ray diffraction is described by the Bragg's law of diffraction (graphical representation in Figure 9). In Bragg's model, each reflection has a set of evenly spaced planes in the crystal, passing through the centers of the atoms of the crystal lattice. These planes are classified with Miller indices (*h,k,l*) (or formally reciprocal lattice vector components) and spacing between the planes d. Bragg proposed that X-rays are scattered 'mirror-like' from the planes. As a consequence of this assumption, the beam scattered from adjacent planes will interfere constructively when the angle θ (between the plane and the beam) will result in path-length difference equal to integer multiple n of the beam wavelength.

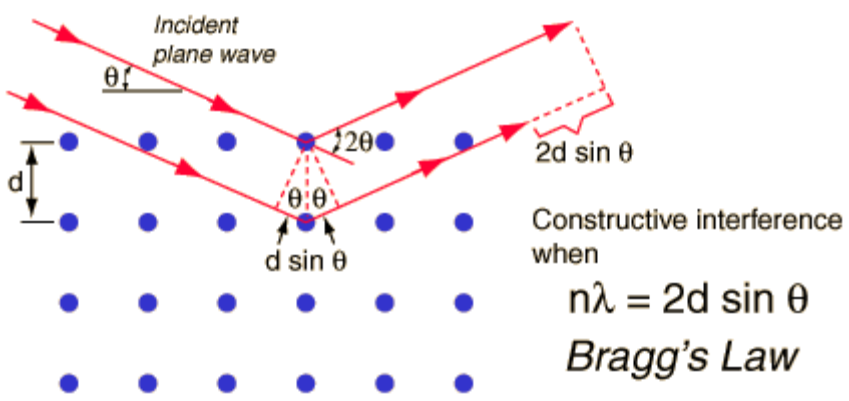


Figure 9. Graphical representation of the Bragg diffraction from a crystal with evenly spaced planes.

$$2d \sin \theta = n\lambda$$

In more practical terms, reflection is indexed when Miller indices are established from known wavelength and scattering angle 2θ. This operation gives unit-cell dimensions, angles, and the space group. To obtain all the data required for structure determination that information has to be complemented by Fourier transform calculations that take into account relative intensities of the reflections. [17][18][19]

A Fourier transform is a mathematical method for decomposing a signal into the frequencies that make it up. In spectroscopic analysis (or any other practical applications) this operation can be reversed to Fourier synthesis of a frequency domain representation that combines the contributions of all the different frequencies to recover the original function of time. This combined with chemical knowledge of a protein is essential in processing the electron density map to get valuable insights into a structure.

From more theoretical approach the electron density $f(r)$ throughout the crystal is assessed, where r corresponds to a three-dimensional position vector within the crystal. To reach this there is a need to collect all the data about its Fourier transform $F(q)$, and then invert it mathematically (Fourier synthesis) in order to access the density defined in a real space. This is done with the formula

$$f(r) = \frac{1}{(2\pi)^3} \int F(q) e^{iq \cdot r} dq$$

where the integral is calculated over all values of q. The vector q in a real space represents a point in reciprocal space, meaning it corresponds to a specific oscillation in electron density as one move in the direction in which q is pointing. The length of q resembles 2π divided by the oscillation wavelength. Then the Fourier transform will be expressed as

$$F(q) = \int f(r)\, e^{-iq \cdot r} dr$$

And here the integral is taken over all possible values of the position vector r in the crystal. The Fourier transform result is normally a complex number, that has a magnitude $|F(q)|$ and phase $\varphi(q)$ with the relation

$$F(q) = |F(q)| e^{i\varphi(q)}$$

X-ray diffraction experiment gives the intensities of the reflections, thus magnitudes $|F(q)|$ but does not provide any information about the phases φ(q). There are certain methods

to overcome this problem and its choice depends on the information available. Molecular Replacement (MR) is one way to go and is used only when a similar structure (threshold of ~30% for protein identity) is known. The model helps in regulating and determining molecule position and orientation. Here, the protein taken into consideration has similar structures characterized and uploaded on PDB, hence molecular replacement was found to be a good solution for the phase problem and allowed for electron density map generation. Other methods for solving the phase problem include:

- Multiple Isomorphous Replacement (MIR) – where new diffracting centers are introduced to the lattice by soaking the crystal in a solution containing heavy atoms (and since they are electron-rich the contribution to the diffraction diagram will be pronounced).
- Single or Multi-wavelength Anomalous Diffraction (SAD or MAD) – these two methods are based on an anomalous scattering of the inner electrons of heavy atoms. Here heavy atoms are considered those that can be naturally present in the protein as a cofactor or an amino acid (such Fe atoms or selenomethionine) and will absorb X-rays at a particular wavelength (absorption edge) and emit them with some delay. This delay causes phase shift that helps to localize them within the structure.

3.4.2 Synchrotron radiation source

X-ray is a type of an electromagnetic radiation, generated in parcels of energy called photons. Its range of wavelength is between 0.01 and 10 nm. They are divided into two different-in-energy waves, first one called soft x-ray, and higher in wavelength called hard x-ray. The energy range of x-rays lays in a range from 100 eV up to 100 keV. From a physical point of view, there are two atomic processes that are responsible for x-rays generation. The first one called white radiation (or Bremsstrahlung radiation), and the second one being K-shell emission. Both of those are related to heavy atoms elements and choice of an element depends on energy/application required. Due to their penetrating nature, they found to be an excellent tool for probing morphology of matter. [20] In scientific research, synchrotron radiation is preferred for an analysis due to its excellent properties.

Synchrotron radiation is a special type of light generation source in wide wavelength range and is considered as one of the most powerful radiation sources due to its collimation and linear polarization. [21] By definition synchrotron is a particle accelerator in which leading magnetic field is time-dependent, and as the name suggest particle beam is traveling in a synchronized path. If particles enriched with high energy are in rapid motion and travel in a curved path (induced by the magnetic field) the synchrotron radiation is generated (as an energy loss in a form of a photon). The scheme of a synchrotron light source is presented in Figure 10. The radiated power is given by the relativistic Larmor formula while the force on the emitting electron is given by the Abraham–Lorentz–Dirac force. The radiation pattern can be distorted from isotropic dipole pattern into an extremely forward-pointing cone of radiation. Synchrotron radiation is the brightest artificial source of X-rays. The planar acceleration geometry appears to make the radiation linearly polarized when observed in the orbital plane, and circularly polarized when observed at a small angle to that plane.

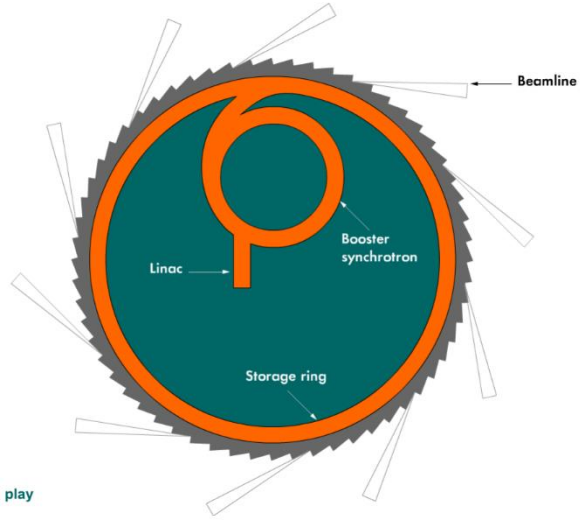


Figure 10. Synchrotron facility scheme.

The magnetic field in the synchrotron is varying in time in order to accelerate particles if almost relativistic speed is reached then the frequency of the electromagnetic field is adjusted to follow their abnormal circulation time. The objective is to adapt parameters in such a way that as particles gain energy, all of them reach the same (thus constant) circulation path. The advantage of such design is vacuum chamber in shape of a thin torus (or ring). The profile of the ring makes more efficient use of a magnetic field. The toroid shape of large scale synchrotrons was found to follow strong focusing principle. [22]

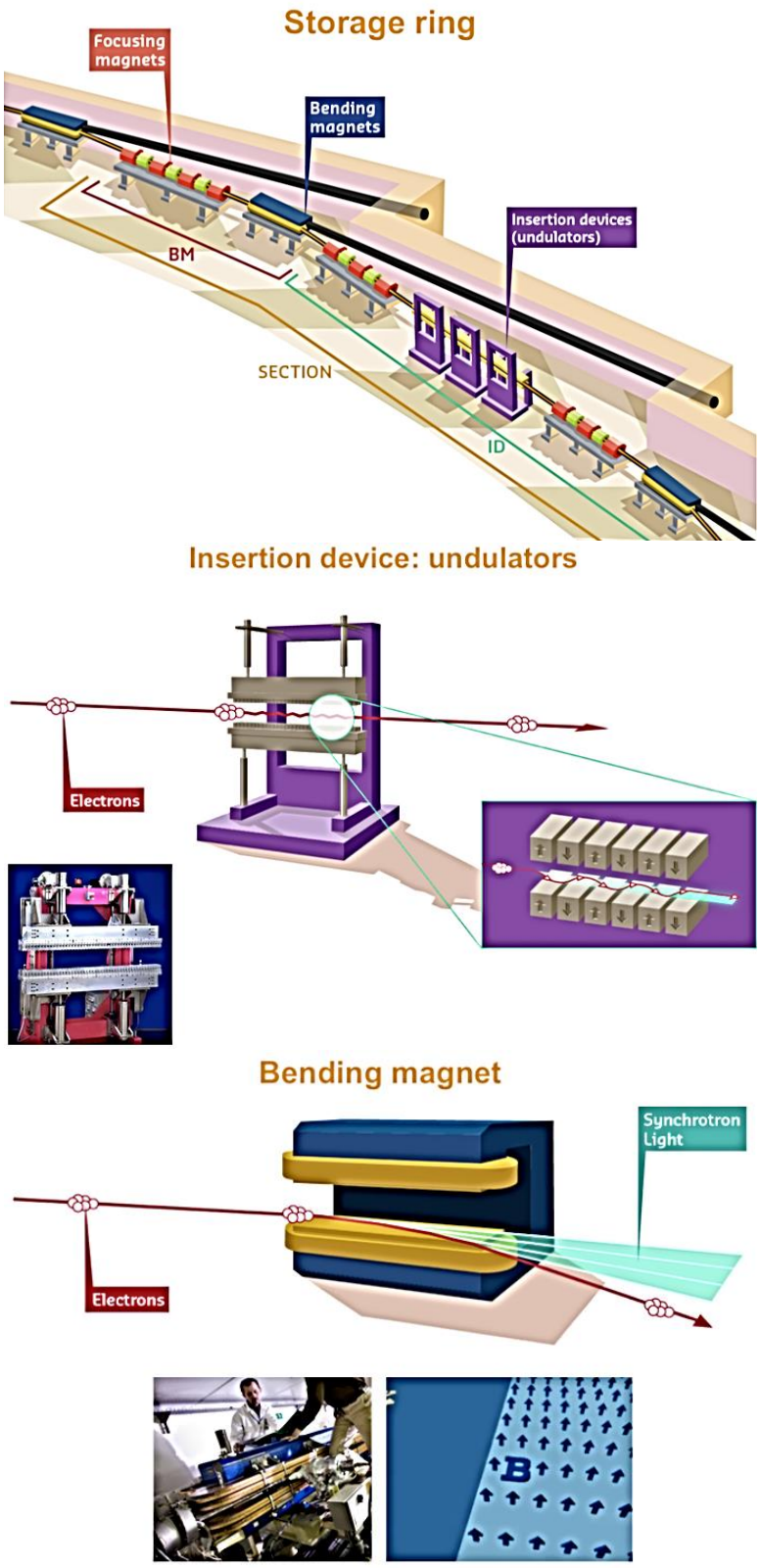


Figure 11. Scheme of the main elements of the storage ring.

The principal structures found at synchrotron are common for every facility of this kind (any details given here are associated to ESRF – European Synchrotron Radiation Facility in Grenoble). First, the storage ring, is an element giving characteristic shape to the facility. It is almost 900 m long tube, where electrons are kept at relativistic speeds and consequently have to be kept at low pressures (~10-9 mbar). As they travel in the ring they pass through different types of magnets (described below) to ultimately generate x-rays.

Booster synchrotron is an element responsible for accelerating electrons before being injected to the storage ring. Having around 300 meters pre-accelerator must give an energy of 6 billion electron volts (6 GeV) to the particles. It is in use only a few times a day for few minutes when storage ring needs to be re-filled. It can inject parcels of 6 GeV electrons every 50 ms.

Linac is an abbreviation for the linear accelerator, and its role to generate electrons for the storage ring in a similar manner to classical cathode ray tubes. Here the electrons reach up to 200 million eV before entering booster synchrotron.

X-ray beams emitted from electrons under magnetic field are directed into beamlines. Beamlines are located in tangent lines to the storage ring. Every beamline is designed for different experiments, thus having different characteristics for shaping the beam and adjusting the wavelength to the proper instrument for analysis.

We can section the storage ring into function-specific devices. Major components of a synchrotron storage ring are:

- Radiofrequency cavities – type of resonator consisting of a closed metal structure that restricts electromagnetic field in the microwave region of the spectrum. In synchrotron used for direct acceleration.
- Dipole magnets (bending magnets) – electromagnet used for the creation of homogeneous magnetic field over some distance. Particles entering dipole magnet will travel in a circular or helical trajectory. In synchrotrons used for deflection of particles.
- Quadrupole magnet (Q-magnet) – as the name suggest it is a group of four magnets fixed in a plane which caused multipole expansion of the field. The advantage of this magnet type is a creation of magnetic field that magnitude grows very rapidly with the radial distance. In synchrotron used for beam focusing.
- Sextupole magnet – the principle is the same as in Q-magnet, but having 6 magnets oriented with alternating north/south pole around an axis.
- Wigglers are a series of magnets designed to periodically and laterally deflect ('wiggle') a beam of charged particles (invariably electrons or positrons) inside a storage ring of a synchrotron. These deflections create a change in acceleration which in turn produces emission of broad synchrotron radiation tangent to the curve, much like that of a bending magnet, but the intensity is higher due to the contribution of many magnetic dipoles in the wiggler.
- Undulators consist of a periodic structure of dipole magnets. The static magnetic field is alternating along the length of the undulator with a wavelength λ_u.

Electrons traversing the periodic magnet structure are forced to undergo oscillations and thus to radiate energy.

The highest, reachable energy is relative to the maximum strength of the magnetic field and minimum radius of the particle path. To increase this energy limit superconducting magnets have been introduced (since they are not limited by so-called magnetic saturation). The cross-section of a storage ring components and the general working principle is shown in Figure 11. [23]

3.4.3 Data collection

XRD experiments for both PROJECT1 and PROJECT2 were performed at ESRF in Grenoble (France) on ID23.1 beamline. This beamline is dedicated for macromolecular crystallography and split into 2 independent workstations. First one (ID23.1) designed for MAD (Multi-wavelength Anomalous Dispersion) measurements in the energy range from 5 up to 20 keV and second (ID23.2) with fixed energy (14.2 keV) and a focused microbeam (8x6 micron).

The characteristics of ID23.1 is the high demagnification ratio (1:4.5) that allows finely focused beam at a crystal position. This is advantageous in terms of protein samples that often give small crystals and even 20 μm crystals can be analyzed. The beam profile has Gaussian shape with FWHM (full width at half maximum) of dimensions 30 μm vertically and 40 μm horizontally with the continuous intensity monitoring system.

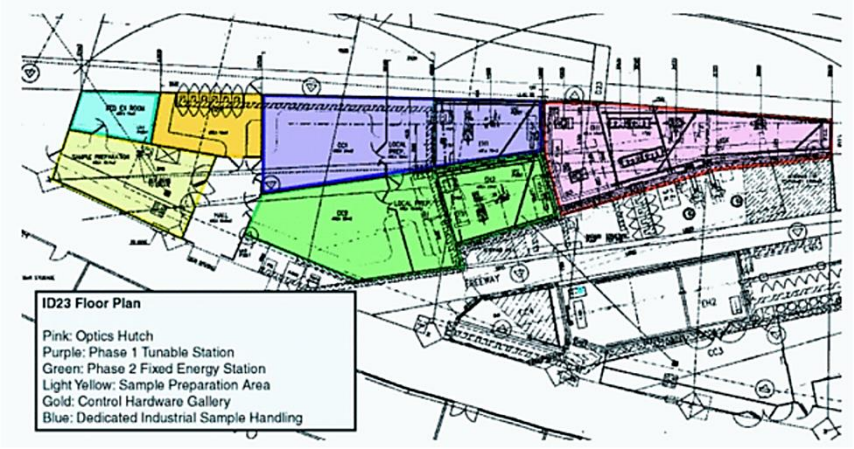


Figure 12. ID23.1 beamline design for Macromolecular crystallography (MX).

The beamline is highly automated granting quick and reliable measurements while keeping it user-friendly. There is sample preparation room (if needed) equipped with optical microscopes. The user benefits from state-of-the-art diffractometers and sample

changer. Remote semi-automatic sample alignment is possible from the control cabins, with experimental parameters calculated automatically based on pre-set user requirements. ID23-1 is equipped with a PILATUS 6M-F which has an active area of 2463 x 2527 pixels and can collect 25 images per second with a readout time of 3 ms. The scheme of the beamline endstation is shown in Figure 12. [23] Diffraction images were recorded at 100 K, with a wavelength of 0.976 Å, an oscillation angle of 0.1°, a sample-to-detector distance of 340.50 mm. A number of 3600 images, corresponding to all sphere, were recorded.

3.4.4 Data analysis and refinement

The data collected at ESRF were saved for further processing with CCP4 software. [24] Firstly set of a diffraction data were chosen according to the best statistics and estimates. These include overall and outer shell parameters, such as resolution, I/σ, and multiplicity. Then *mtz*, *pdb* and *fasta* sequence files were collected for CCP4 software processing and model building. For phasing and molecular replacement the structure of 40% identity was taken (PDB ID: 1DNA) according to the sequence alignment with Clustal Omega. After model building, amino-acid residues were manually substituted and fit into electron density map with Coot software [25].

The data at a resolution of 1.713 Å for GR1 were processed using XDS [26], scaled with SCALA [27] and TRUNCATE [28]. The correct space group was determined with POINTLESS [29] and confirmed in the structure solution stage. The data collection statistics are reported in the discussion section.

The GR1 structure was solved by molecular replacement using the program MOLREP from CCP4 package. The coordinates of GR from *Saccharomyces cerevisiae* (PDB code 2HQM) [30], deprived of chain B, FAD and water molecules, were used as search probe. The residual electron density map clearly showed the position of FAD, which was added to the model. Water molecules were automatically added and, after a visual inspection, they were conserved in the model only if contoured at 1.0 σ on the ($2F_o - F_c$) map and if they fell into an appropriate hydrogen bonding environment. All stages of the refinement were performed with REFMAC 5.5.0109 [31], alternate conformations, side chains positions if visible, were inserted/modified. The manual rebuilding was performed with

Coot. [32] There is also possibility to align subunits (or a whole enzyme) with use of LSQKAB software within CCP4 software package in order to visualize overlapping of the secondary structure elements of two different chains (for instance, GRs from two different organisms).

The crystallographer deposits a model of the asymmetric unit of the crystal in the PDB, along with the experimental diffraction data (amplitudes and widths of the X-ray reflection spots, or structure factors) from which the electron density map can be reconstructed. Examining the relation between the published model and the electron density map provides clear insight into the uncertainties in the model than does merely examining the model itself. In addition to examining the entire map ($2F_0$-F_c) it is revealing to examine the difference map (F_0-F_c), which shows where the model fails to account for the map.

Refinement is a continuous process of adjustment of the model to maximize the agreement of the structure factors of the model with the structure factors experimentally observed. For GR structure twelve cycles of refinement were performed with *Refmac* and with gradually decreasing R value. This agreement is mathematically expressed as

$$R = \frac{\sum_{hkl} \left| |F_{hkl}^{oss}| - |F_{hkl}^{calc}| \right|}{\sum_{hkl} |F_{hkl}^{oss}|} \times 100$$

And generally its value lays in the range between 0.15 – 0.25 for well resolved structures (here 0.24).

Chapter 4 - Results and discussion

Although a number of GR structures are found in PDB there is little known about plant origin isoforms. Sequence alignment shown low identity with bacterial, human and yeast isoforms, nevertheless certain structural features, like catalytic cysteines conservation. Yet it is important to investigate elementary processes like cell homeostasis in all kingdoms and plant one seems to be suffering from the little attention paid to structure-function relationships. Exposure to a drastically different environment and contrasting cell cycle makes plant proteins interesting objects for structural studies in terms of evolutionary biology.

4.1 Protein sample quality

Biocrystallography requires high-quality material for repetitive crystallization experiments. Purification is crucial in this case since it will directly affect the rate, quality and likely morphology of the protein crystals. Glutathione reductase 1 and glutathione reductase 2 (GR1 and GR2) were expressed in *E. Coli* and purified to homogeneity by Ni2+ affinity chromatography. The yield of GR1 and GR2 products was 15 and 10 mg/L respectively. The GR1 sequence has 501 amino acid residues (Mw=53,128 Da) and GR2 has 488 residues (Mw=52,124 Da), including the histidine tag required for purification (His-tag peptide: MHHHHHHTM). The sequence of GR1 isoform is found in Uniprot database under the code of A8HXA6. The identity between GR1 and GR2 is little over 50% and pairwise alignment is presented below.

```
GR1_cytoplasmic       -------MAEEYDLVTLGAGSGGVRASRFAATLYGAKVACVELPFGFVSSETVGGAGGTC
GR2_chloroplastic     MASASAGEQFDFDLFTIGAGSGGVRGSRFAS-SYGAKVAVCELPFDYISSDTKGGVGGTC
                                ::**.*:********.****:  ******  ****.::**:* **.****

GR1_cytoplasmic       VIRGCVPKKLLVYGAAYAEEFADARGFGWALPAAGAGAEGGPAHDWASLMKLKEKEITRL
GR2_chloroplastic     VLRGCVPKKLMVYASEYAEEFKASQGFGWTLPP------GAATHSWTAFIEAKRKELQRL
                      *:********:**.: *****  ::****:**       *. :*.*::::: *.**: **


GR1_cytoplasmic       NSTYGNILKNANVALIEGRGALKDAHTVEVTAADGSVRLLKAKHVLIATGGVATAIPMEG
GR2_chloroplastic     NGAYKNTLKNAKVELVEGRGRVVDAHTVEVDGK-----RYRAKNILIAVGGKPHKLDIPG
                      *.:* * ****:* *:**** : ******* .        :**::***.**    : : *

GR1_cytoplasmic       AEHAIMSDDALALQSLPPGPIVVLGAGYIATEFAGIFRGTHAAQYAVHLMFRGDKVLRGF
GR2_chloroplastic     AELCITSDEALELPACP-QKVAVLGGGYIAVEFAGIFSRFGA---EVHTVYRQPLPLRGF
                      ** .* **:** * : *   :.***.****.******    *    ** ::*    ****
```

```
GR1_cytoplasmic         DEECRDQVQDNLTRRGIHLHPGCKPTKLEKHGEGDLTLHYTDGTGAAQSL-KCGLVMMAT
GR2_chloroplastic       DEEVRKFAAEQYAAAGLHLHAGCNPVSVSKQPNGKLSLVVKGPDGATSTLTDLDQVMMAT
                        *** *. . :: :  *:*** **:*..:.*: :*.*:*  ..  **:.:* . . *****

GR1_cytoplasmic         RRKPRVDGIGLEAVGVALDGQGAIKVDEFSRTNVPDVWAIGDVTNRINLTPVALMEGMAF
GR2_chloroplastic       GRVPKTSGLGLEEAGVKMGSKGQVLVDEYCRTNVPSIWAVGDVIDRIQLTPVALMEGMAV
                         * *:..*:*** .** :..:* : ***:.*****.:**:*** :**:***********.

GR1_cytoplasmic         AKSCFGGELTKPDYRNVASAVFCQPPLATVGYTEEQAVKEFAGNIDVYVSRFRPMKYTIS
GR2_chloroplastic       AKSVAHNTPTVPDYFAVPSAVFSNPEIATVGYSEEQAAEKY-GDVDIYTTSFKPMRNTIS
                        ***   .  * ***  * ****.:* :*****:****.::: *::*:*.: *:**: ***

GR1_cytoplasmic         GREEKTLMKLIVHAESDVVLGCHMVGPDAPEIMQGLAVALKCGATKAQFDSTVGIHPTAA
GR2_chloroplastic       GSPIRTFMKIVVDAASQKVVGMHMVGAEAAEIMQGFAVAVKVGATKQQLDSVVGIHPSAA
                        *   :*:**::*.* *: *:* **** :* *****:***:* **** *:**.*****:**

GR1_cytoplasmic         EEFVTMRSRSRRVPATGTSKL
GR2_chloroplastic       EEFVTMRTVSRQVRKPA----
                        *******: **:*   .
```

Sequence alignment between two GR isoforms from *Chlamydomonas reinhardtii*. The identical residues are marked with "*****", while similar amino acids are marked with ":". The alignment performed with Clustal Omega online tool available on EMBL-EBI website.

The molecular weight of a single protein chain was assessed by SDS-PAGE exhibiting bands around 50 kDa. The same electrophoretic profile was recognized for oxidized forms of GRs indicating a lack of intermolecular disulfide bridge(s) – indicated respectively by "-" for reducing and "+" for oxidized condition (Figure 13A).

The quaternary structure of the proteins was estimated with Dynamic Light Scattering (DLS) experiment. This non-destructive technique gives immediate access to size distribution evaluation of small particles (not bigger than 1 μm) in solution as described in *Methods and experimental fine points* section. Since it is known that the biological assembly of GRs is a homodimer the expected molecular weight calculated from the hydration radius (R_h) would be the double of a single protein chain, thus around 100 kDa. The mean R_h based on the values from 5 independent measurements, resulted in a good accordance with the molecular weight of the dimer. Indeed, the R_h of GR1 was 5.17 ± 0.19 nm, corresponding to an apparent molecular weight of 122 ± 2 kDa, and the R_h value determined for GR2 was 4.94 ± 0.22 nm, a value corresponding to an apparent molecular mass of 112 ± 3 kDa. Exemplary plot of size distribution by volume is shown in Figure 13B.

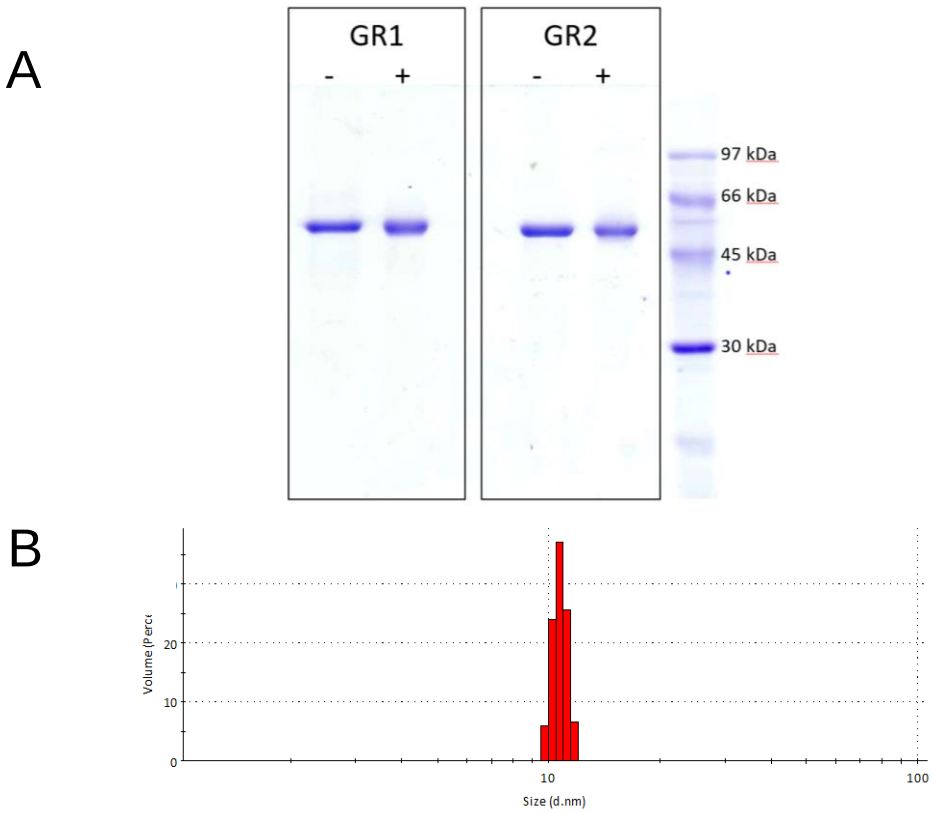


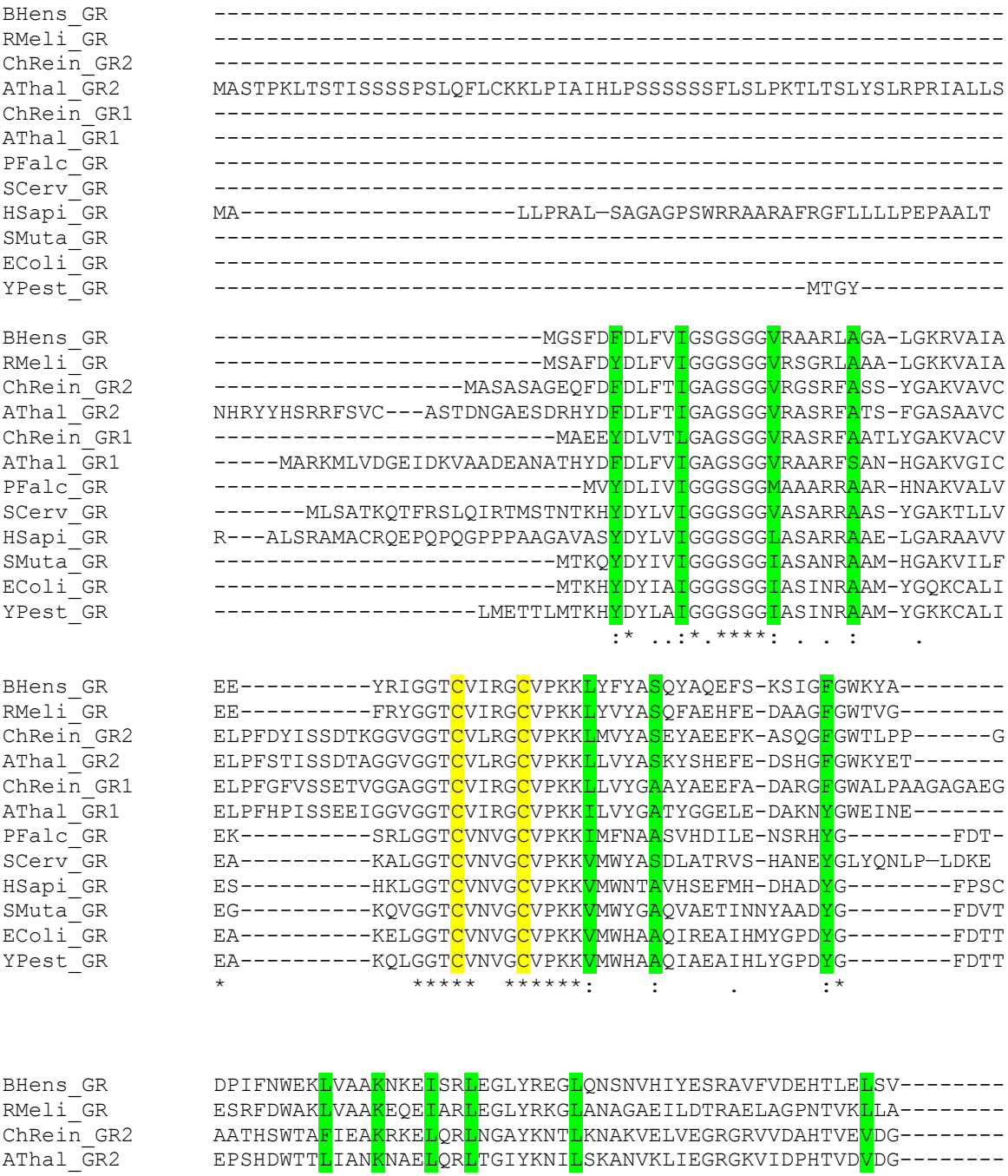
Figure 13. (A) SDS-PAGE of GR1 and GR2."+" indicates oxidized form, "-" indicates reduced enzyme. (B) DLS analysis GR1 and GR2 reported as size distribution by volume in function of hydrodynamic diameter.

4.2 Sequence analysis

Protein required for a single and manual crystallization trial is about 1 to 5 µL, but when considering the range of possible conditions to grow an appropriate crystal, a number of trials have to be set up and thus larger volume is required. To limit this to some extent it is convenient to make an educated guess based on existing crystal structures and analyzing the protocols for crystallization. Indeed for GR, there are many crystal structures available on PDB from various organisms. The possible conditions were collected and identity evaluated. On the next page is shown sequence alignment for eight existing GR structures from human, few bacteria, yeast, protista with plant GR1. Transferases and glutathione amide reductases have not be considered in this comparative analysis.

Multiple sequence alignment of GRs from different organisms was done with Clustal Omega from EBI online tools. The first code is an abbreviation of the organism name.

Chlamydomonas reinhardtii - ChRein, *Homo sapiens* - HSapi, *Plasmodium falciparum* - PFalc, *Saccharomyces cerevisiae* - SCerv, *Streptococcus mutans* - SMuta, *Escherichia coli* - EColi, *Yersinia pestis* - YPest, *Bartonella henselae* - BHens, *Rhizobium meliloti* - RMeli, *Arabidopsis thaliana* - AThal. In yellow catalytic residues are highlighted, hence cysteines, tyrosines, and histidines (while histidines are belonging the second subunit of an enzyme). Green markings indicate conservative mutations.



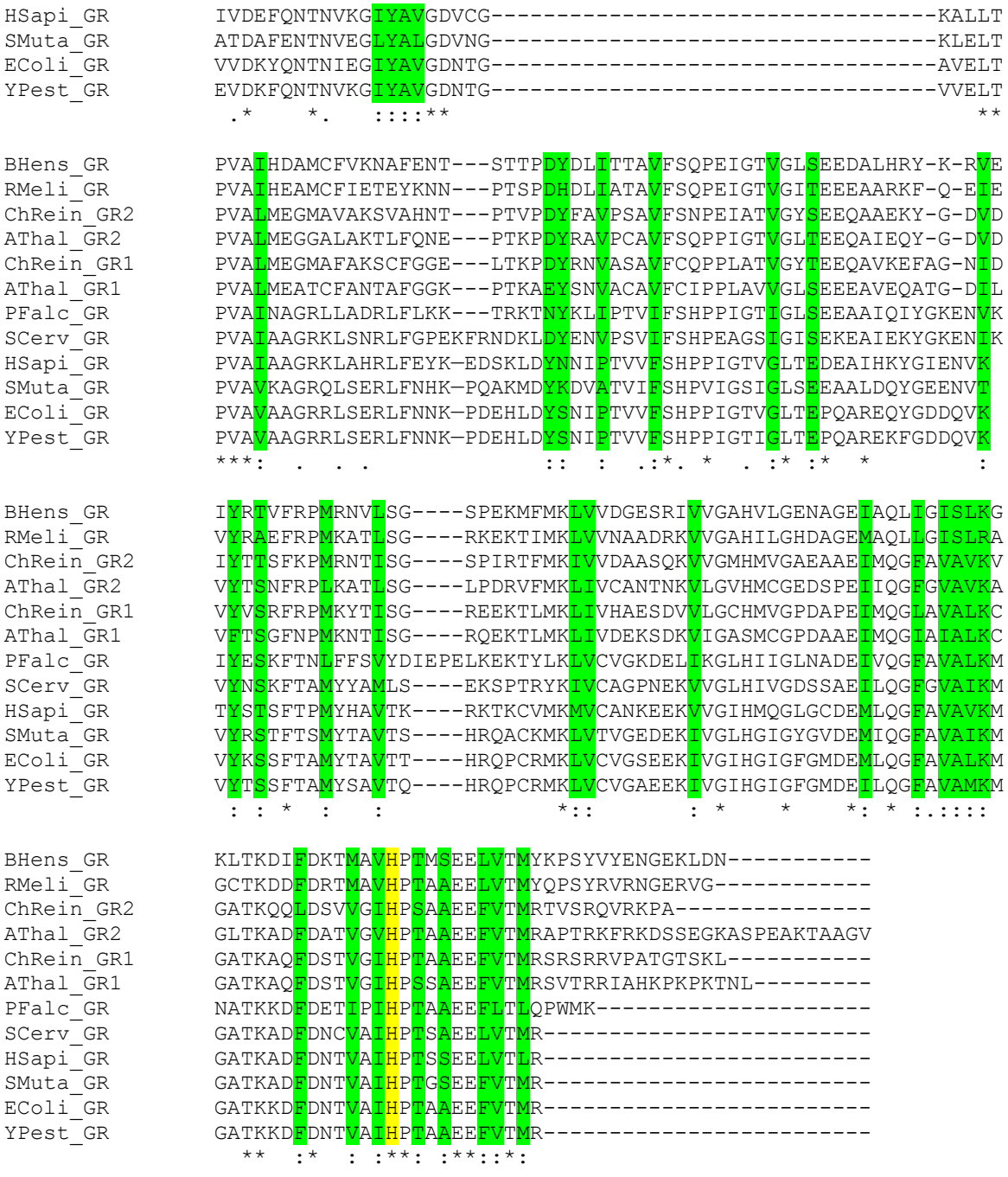
```
ChRein_GR1      GPAHDWASLMKLKEKEITRLNSTYGNILKNANVALIEGRGALKDAHTVEVTA--------
AThal_GR1       KVDFTWKKLLQKKTDEILRLNNIYKRLLANAAVKLYEGEGRVVGPNEVEVRQ--------
PFalc_GR        KFSFNLPLLVERRDKYIQRLNNIYRQNLSKDKVDLYEGTASFLSENRILIKGTKDNNNKD
SCerv_GR        HLTFNWPEFKQKRDAYVHRLNGIYQKNLEKEKVDVVFGWARFNKDGNVEVQK--------
HSapi_GR        EGKFNWRVIKEKRDAYVSRLNAIYQNNLTKSHIEIIRGHAAFTSDPKPTI----------
SMuta_GR        TQAFHFDVLKQNRQAYIDRIHDSYERGFDSNGVERVYGYATFVDA—HTV----------
EColi_GR        INKFNWETLIASRTAYIDRIHTSYENVLGKNNVDVIKGFARFVDA—KTL----------
YPest_GR        VNHFDWKKLIANRTAYIDRIHQSYERGLGNNKVDVIQGFARFVDA—HTV----------
                   .    :   :   : *:   * . : .         . .       :

BHens_GR        -T---GERISAEKILIATGAKIVSNS---AIKGSDLCLTSNEIFDLEKLPKSI-VIVGGG
RMeli_GR        -S---GKTVTAERIVIAVGGHPSPHD---ALPGHELCITSNEAFDLPALPESI-LIAGGG
ChRein_GR2      ------KRYRAKNILIAVGGKPHK-L---DIPGAELCITSDEALELPACPQKV-AVLGGG
AThal_GR2       ------KIYTTRNILIAVGGRPFI-P---DIPGKEFAIDSDAALDLPSKPKKI-AIVGGG
ChRein_GR1      -ADGSVRLLKAKHVLIATGGVATA-I---PMEGAEHAIMSDDALALQSLPPGPIVVLGAG
AThal_GR1       -IDGTKISYTAKHILIATGSRAQK-P---NIPGHELAITSDEALSLEEFPKRA-IVLGGG
PFalc_GR        NGPLNEEILEGRNILIAVGNKPVF-P---PVKGIENTISSDEFFNIKES-KKI-GIVGSG
SCerv_GR        -RDNTTEVYSANHILVATGGKAIF-PE—NIPGFELGTDSDGFFRLEEQPKKV-VVVGAG
HSapi_GR        --EVSGKKYTAPHILIATGGMPST-PHESQIPGASLGITSDGFFQLEELPGRS-VIVGAG
SMuta_GR        --EVAGEHYTAPHILIATGGHALL-P---DIPGSEYGITSDGFFELDAIPKRT-AVVGAG
EColi_GR        --EVNGETITADHILIATGGRPSH-P---DIPGVEYGIDSDGFFALPALPERV-AVVGAG
YPest_GR        --EVNGETITADHILIATGGRPSH-P---DIPGAEYGIDSDGFFELDEMPKRV-AVVGAG
                             .:::*.*           : * .    *:  : :         : .

BHens_GR        YIGVEFANIFHGLGV---KTTLLHRGDLILRNFDYDLRQLLNDAMVAKGISIIYEATVSQ
RMeli_GR        YIAVEFANIFHGLGV---KTTLIYRGKEILSRFDQDMRRGLHAAMEEKGIRILCEDIIQS
ChRein_GR2      YIAVEFAGIFSRFGA---EVHTVYRQPLPLRGFDEEVRKFAAEQYAAAGLHLHAGCNPVS
AThal_GR2       YIALEFAGIFNGLNC---EVHVFIRQKKVLRGFDEDVRDFVGEQMSLRGIEFHTEESPEA
ChRein_GR1      YIATEFAGIFRGTHAAQYAVHLMFRGDKVLRGFDEECRDQVQDNLTRRGIHLHPGCKPTK
AThal_GR1       YIAVEFASIWRGMGA---TVDLFFRKELPLRGFDDEMRALVARNLEGRGVNLHPQTSLTQ
PFalc_GR        YIAVELINVIKRLGI---DSYIFARGNRILRKFDESVINVLENDMKKNNINIVTFADVVE
SCerv_GR        YIGIELAGVFHGLGS---ETHLVIRGETVLRKFDECIQNTITDHYVKEGINVHKLSKIVK
HSapi_GR        YIAVEMAGILSALGS---KTSLMIRHDKVLRSFDSMISTNCTEELENAGVEVLKFSQVKE
SMuta_GR        YIAVEISGILHALGS---ETHLFVRRDRPLRKFDKEIVGTLVDEMKKDGPHLHTFSVPKE
EColi_GR        YIAVELAGVINGLGA---KTHLFVRKHAPLRSFDPMISETLVEVMNAEGPQLHTNAIPKA
YPest_GR        YIAVEIAGVLNGLGT---ETHLFVRKHAPLRTFDPLIVETLLEVMNTEGPKLHTESVPKA
                **. *: .:             . *    *  **              .  .

BHens_GR        VQSTENCYNVV-L--------TNGQTICADRVMLATGRVPNTTGLGLERAGVKVNEFGAV
RMeli_GR        VSADADGRRVATT--------MKHGEIVADQVMLALGRMPNTNGLGLEAAGVRTNELGAI
ChRein_GR2      VSKQPNGKLSLVVKGPDG---ATSTLTDLDQVMMATGRVPKTSGLGLEEAGVKMGSKGQV
AThal_GR2       IIKAGDGSFSLKT--------SKGTVEGFSHVMFATGRKPNTKNLGLENVGVKMAKNGAI
ChRein_GR1      LEKHGEGDLTLHYTDGTG---AAQ-SLKCGLVMMATRRKPRVDGIGLEAVGVALDGQGAI
AThal_GR1       LTKTDQGIKVISS--------HGE-EFVADVVLFATGRSPNTKRLNLEAVGVELDQAGAV
PFalc_GR        IKKVSDKNLSIH-------LSDGRIYEHFDHVIYCVGRSPDTENLKLEKLNVETN-NNYI
SCerv_GR        VEKNVETDKLKI------HMNDSKSIDDVDELIWTIGRKSH-LGMGSENVGIKLNSHDQI
HSapi_GR        VKKTLSGLEVSMVTAVPGRLPVMTMIPDVDCLLWAIGRVPNTKDLSLNKLGIQTDDKGHI
SMuta_GR        VIKNTDNSLTLIL--------ENGEEYTVDTLIWAIGRAANTKGFNLEVTGVTLDSRGFI
EColi_GR        VVKNTDGSLTLEL--------EDGRSETVDCLIWAIGREPANDNINLEAAGVKTNEKGYI
YPest_GR        VIKNADGSLTLQL--------ENGTEVTVDHLIWAIGREPATDNLNLSVTGVKTNDKGYI
                :    .                        . ::    *      :  .  .:     . :

BHens_GR        VVDEKMTTNVSHIWAVGDVTG----------------------------------HIQLT
RMeli_GR        IVDAFSRTSTPGIYALGDVTD----------------------------------RVQLT
ChRein_GR2      LVDEYCRTNVPSIWAVGDVID----------------------------------RIQLT
AThal_GR2       EVDEYSQTSVPSIWAVGDVTD----------------------------------RINLT
ChRein_GR1      KVDEFSRTNVPDVWAIGDVTN----------------------------------RINLT
AThal_GR1       KVDEYSRTNIPSIWAVGDATN----------------------------------RINLT
PFalc_GR        VVDENQRTSVNNIYAVGDCCMVKKSKEIEDLNLLKLYNEERYLNKKENVTEDIFYNVQLT
SCerv_GR        IADEYQNTNVPNIYSLGDVVG----------------------------------KVELT
```



A colon (:) in a sequence alignment indicates conservative mutations meaning that even if the exact amino acid is not conserved the biochemical properties stays similar (like hydrophobicity, charge or size). An asterisk (*) indicates conserved sequence. For various GRs, quite a few primary structure characteristics are sustained, such as cysteines, glycines, serines, prolines, etc. In an alignment, there are few peptides maintaining the same order being likely binding pockets for nucleotides. A single dot (.) is a semi-

conservative mutation and no sign simply means non-conservative mutation. As indicated above conservative mutations are rather abundant and the majority has strongly hydrophobic character. Protein identity with GR1 and accession codes is shown in Table 2.

Table 2. Summary of GR identity with different origin enzymes that are available on *Uniprot* and/or PDB database. In green is highlighted best identity of available structure, and in yellow the identity with homologous plant organism (while the structure is not known).

Organism	Uniprot ID	PDB ID	Identity GR1	Identity GR2
Chlamydomonas reinhardtii (GR1)	A8HXA6	-	-	51.497%
Homo sapiens	P00390	3DK9	29.359%	24.064%
Plasmodium falciparum	Q94655	1ONF	26.802%	18.761%
Saccharomyces cerevisiae	P41921	2HQM	32.819%	24.539%
Streptococcus mutans	Q8DUR5	5V36	34.073%	35.331%
Escherichia coli	P06715	1GER	34.879%	24.811%
Yersinia pestis	Q8CZL0	5VDN	33.992%	39.095%
Bartonella henselae	A0A0H3LWY9	3O0H	36.97%	28.729%
Rhizobium meliloti	Q92PC0	4DNA	40.202%	28.281%
Arabidopsis thaliana (GR1)	P48641	-	50.969%	47.431%
Arabidopsis thaliana (GR2)	P42770	-	43.441%	49.037%

Not surprisingly the highest identity is observed for those belonging to the plant kingdom (*Arabidopsis thaliana*) with values around 50%. In the case of those GRs that are accessible on PDB, identity is lowered to ~40%. In fact, those with known structures are helpful in phase problem solving with molecular replacement. Since the threshold limit for MR is ~30% sequence identity there are technically six possible model structures.

4.3 Crystallization trials and optimization

Once the pure and homogenous sample of protein was obtained crystallization trials were performed. Both isoforms were at concentration of 10 mg/mL in 30 mM Tris-HCl (pH 7.9) and 1 mM EDTA. The screening kit from Molecular Dimensions Structure Screen 1 cacodylate free MD1-01-CF was utilized and for GR1 all 50 conditions were tested while due to volume limitation of GR2 only 24 conditions were taken into account. Conditions for GR2 were picked as an educated guess from the existing crystallization conditions for GR from various organisms.

Plates were kept in the protein incubator at 20 ± 1 °C and regularly checked under the polarized microscope every seven days. After two weeks few positive results were observed for GR1, with varying size and morphology of the crystals (Figure 14). In the case of GR2, the majority of the drops remained clear for about two months, indicating that the starting protein concentration was not sufficient to reach the supersaturation conditions. The most probable explanation is an inaccurate determination of protein concentration after the purification procedure. The only crystals and spherulites grown are shown in Figure 15.

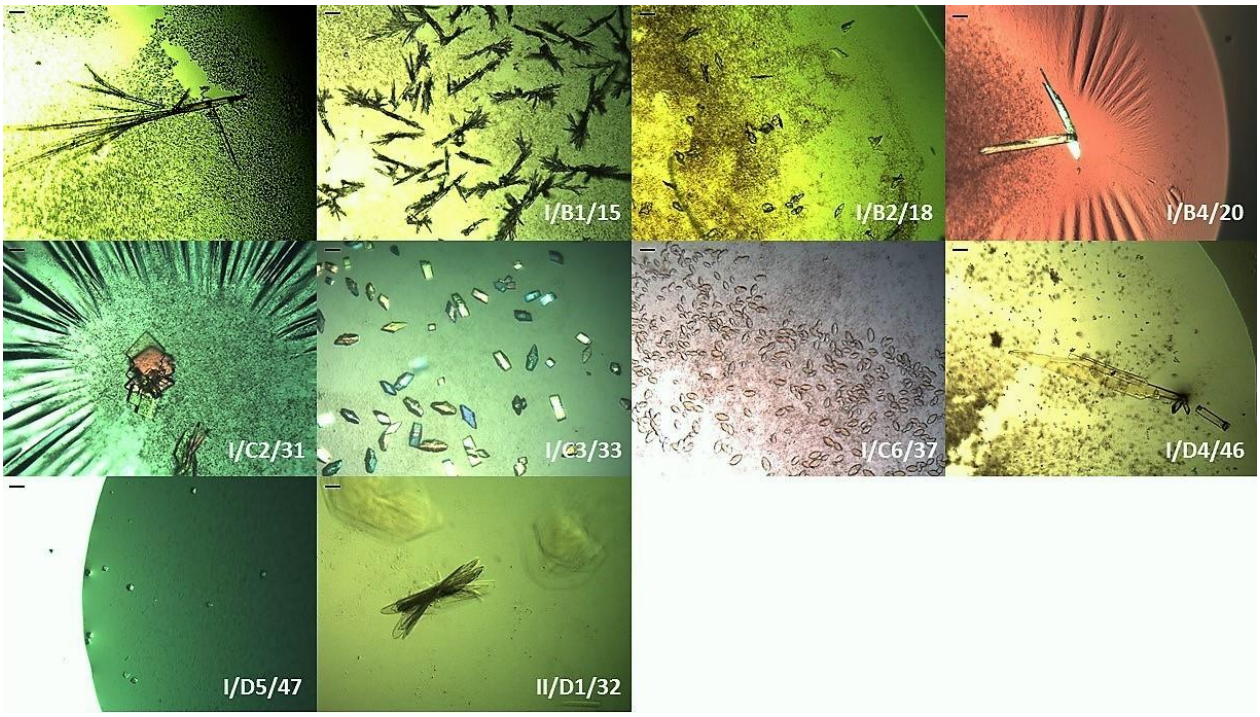


Figure 14. Images of positives hints from crystallization trials for GR1. All images taken at zoom 4x under polarized light. Naming scheme: plate#/well/reagent solution. 70µm scale bar located at every top left corner.

Figure 15. Images of positives hints for GR2. Due to the size and morphology of the crystals, they were not considered in further analysis. Labels and scale bar as in Figure 14.

Crystallization trials gave 10 positive results after only 2 weeks. With this number, it was possible to efficiently design the optimization step. Most promising results were observed for I/C3/33 giving single crystals with well-defined faces, quite large variety of morphologies, and uniform size distribution (around 100 μm each). A similar observation was valid for I/C6/37 in terms of size distribution. Several conditions gave quite large crystal aggregates as I/B4/20, I/C2/31, and I/C2/31, while from conditions I/A4/7 and I/B1/15 branched crystalline aggregates grew. The list of the reservoir solution composition is reported in Table 3.

Table 3. List of reagent solutions for crystallization trials that gave positive results. In green are marked reservoir solutions given priority in optimization step.

N°	Image code	Salt	Buffer (pH)	Precipitant
1	I/A4/7	0.2 M ammonium acetate	0.1 M tri-sodium citrate (5.6)	30% w/v PEG 4 kDa
2	I/B1/15	0.2 M magnesium acetate	0.1 M MES (6.5)	20% w/v PEG 8 kDa
3	I/B2/18	0.2 M sodium acetate	0.1 M MES (6.5)	30% w/v PEG 8 kDa
4	I/B4/20	0.2 M calcium acetate	0.1 M MES (6.5)	18% w/v PEG 8 kDa
5	I/C2/31	None	0.1 M Na HEPES (7.5)	10% v/v 2-propanol, 20% w/v PEG 4 kDa
6	I/C3/33	0.2 M magnesium chloride	0.1 M Tris (8.5)	30% w/v PEG 4 kDa
7	I/C6/37	0.2 M sodium acetate	0.1 M Tris (8.5)	30% w/v PEG 4 kDa
8	I/D4/46	0.05 M potassium dihydrogen phosphate	None	20% w/v PEG 8 kDa
9	I/D5/47	None	None	30% w/v PEG 8 kDa
10	II/D1/32	None	0.1 M Tris (8.5)	2.0 M ammonium sulfate

The table indicates that the best precipitant for GR1 is a polymer as PEG with a molecular weight of 4 or 8 kDa. Improved morphology and size distribution can be observed at a smoothly elevated pH (8.5 for N° 6 and 7) using Tris buffer and high concentration of PEG 4 kDa. When inspecting a salt present in the solution no visible trend is observed since even in cases lacking salt (N° 5, 9 and 10) crystals were formed. Also, N° 3 and 9 were considered to be encouraging targets for crystal form improvement.

For the optimization step, 68 conditions were primarily planned with protein at the same concentration as for the trials. For these experiments, several conditions from two extra

kits were employed (Crystal Extension Kit and JCGS-plus, reagent composition list in additional information) with 20 solutions prepared manually. Main parameters that were adjusted are:

- Precipitant concentration and its molecular weight - Polyethylene glycol with a molecular weight between 3.350 and 8 kDa in concentration range 18 – 30% w/v was utilized;
- pH – certain conditions with lower pH were taken for optimization to see the influence of those below 8.5 (here 6.5 and 7.5);
- Salts – no striking conclusions were withdrawn here, thus most commonly used salts were tested (like magnesium chloride, ammonium sulfate, sodium/calcium/ammonium acetate or none);

Three weeks were required to observe a positive results. Optimization step can be considered successful. As seen in Figure 16 there are conditions where modifications applied lead to the growth of single, clean-face crystals with dimensions suitable for an XRD experiment as for condition III/C6/33B in which a unique single crystal with dimension 270 μm in every direction, was observed. Thus, conclusions withdrawn about precipitant concentration and molecular weight were correct. Table 4 contains reagent solution composition for positives indicating that magnesium chloride, Tris buffer at pH 8.5 and a decrease in PEG 4 kDa concentration (from 30% w/v down to 20% w/v) was the best crystallization condition. Also, condition with 25% w/v PEG 4 kDa (III/C6/33A, Figure 16) resulted in clean morphology crystals of lower dimensions.

Even though conditions N° 7 and 8 gave the best result, more crystals were fished in the case of unexpected events (such as weakly diffracting crystals). Precisely 29 crystals were taken for an XRD experiment at synchrotron radiation source. Before freezing in liquid nitrogen, each crystal was fished with a loop of an appropriate size and immersed in a solution containing the original precipitant concentration plus 20% w/v PEG 400 Da used as a cryoprotectant.

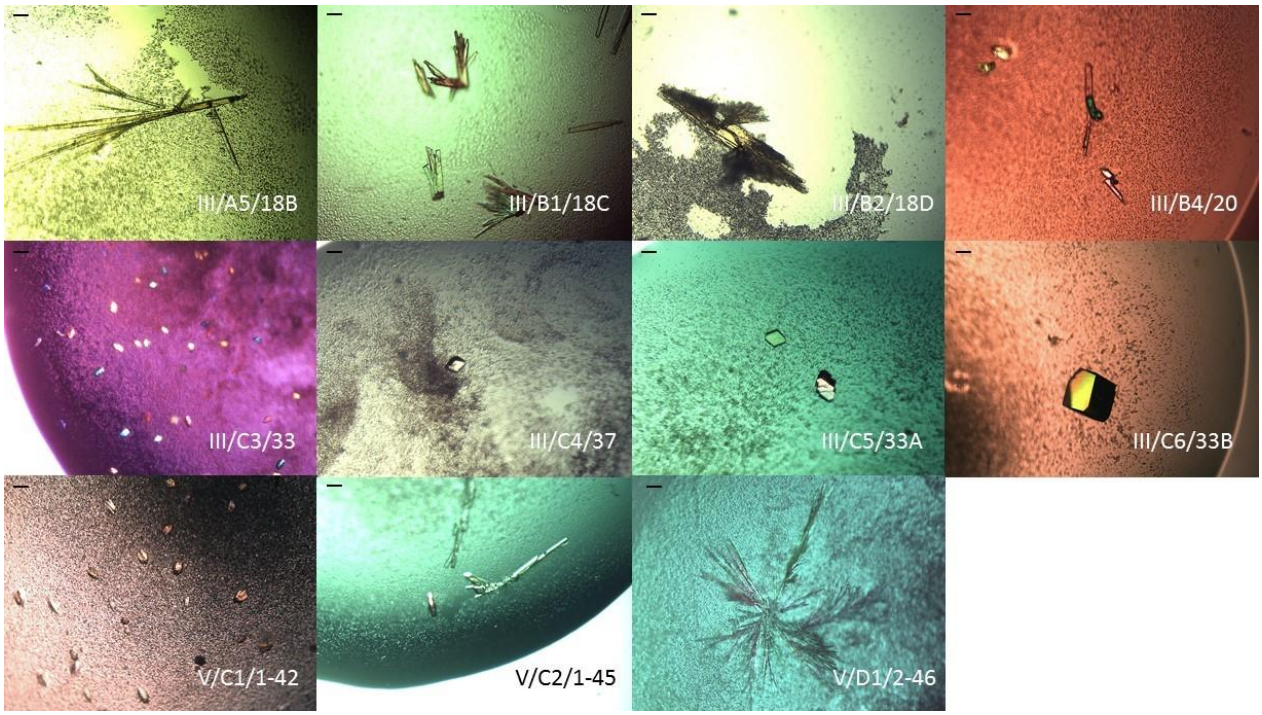


Figure 16. Images of positives from optimization step for GR1. All images taken at zoom 4x under polarized light. Naming scheme: plate#/well/reagent solution. 70 µm scale bar located at every top left corner.

Table 4. List of reagent solutions for optimization step that gave positive results. In yellow marked reagent solutions that gave suitable crystals for an XRD analysis.

N°	Image code	Salt	Buffer (pH)	Precipitant
1	III/A5/18B	0.2 M sodium acetate	0.1 M MES (6.5)	18% w/v PEG 4 kDa
2	III/B1/18C	0.2 M magnesium chloride	0.1 M MES (6.5)	18% w/v PEG 8 kDa
3	III/B2/18D	0.2 M magnesium chloride	0.1 M MES (6.5)	18% w/v PEG 4 kDa
4	III/B4/20	0.2 M calcium acetate	0.1 M MES (6.5)	18% w/v PEG 8 kDa
5	III/C3/33	0.2 M magnesium chloride	0.1 M Tris (8.5)	30% w/v PEG 4 kDa
6	III/C4/37	0.2 M sodium acetate	0.1 M Tris (8.5)	30% w/v PEG 4 kDa
7	III/C5/33A	0.2 M magnesium chloride	0.1 M Tris (8.5)	25% w/v PEG 4 kDa
8	III/C6/33B	0.2 M magnesium chloride	0.1 M Tris (8.5)	20% w/v PEG 4 kDa
9	V/C1/1-42	0.2 M magnesium chloride hexahydrate	0.1 M Tris (8.5)	20% w/v PEG 8 kDa
10	V/C2/1-45	0.17 M ammonium sulfate	None	25.5% w/v PEG 4 kDa, 15% w/v glicerol
11	V/D1/2-46	0.2 M ammonium acetate	0.1 M BIS-Tris (5.5)	25% w/v PEG 3.35 kDa

4.4 Data collection, processing and model building

4.4.1 Data collection

Out of 29 crystals prepared for XRD analysis at ESRF, for 11 crystals full data set was collected. Other crystals were not taken for full data collection due to centering issues, weak diffraction or simply time limitation. Nevertheless, software for data collection at ID23.1 beamline gives a primary estimate about the resolution or more generally the quality of the data based on diffraction pattern. The best resolution of 1.713 Å and 2.109 Å was observed for III/C6/33B crystal and III/C5/33A respectively. The image of a centered and mounted crystal with an exemplary diffraction pattern is shown in Figure 17.

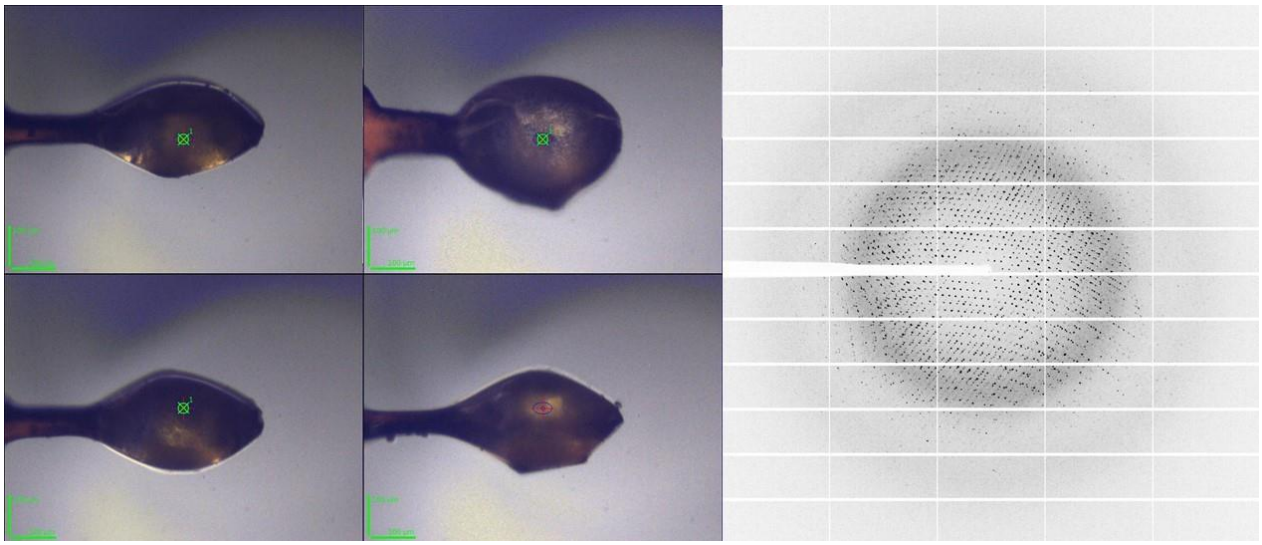


Figure 17. Left: images of III/C6/33B crystal manually centered on a goniometer. Right: exemplary diffraction pattern of a GR1 crystal taken for a full data collection.

Experimental details from the data collection and processing for III/C6/33B sample are summarized in Table 5. Preliminarily processed data is available right after the diffraction experiment and is accessible on ISPyB website (Information System for Protein crystallography Beamlines for ESRF, SOLEIL, MAXIV and few EMBL beamlines). Some data processing is fully automatic, hence indexing (assigning Miller indices for each reflection and the Bravais lattice, determining the unit cell parameters), merging and scaling (to merge the reflections related to the same symmetry operations and assigning the space group) and structure factor (F_{hkl}) that is calculated from the intensities of a diffraction spot.

Table 5. Summary of data collection, refinement and Ramachandran plot details for the III/C3/33B sample.

Parameter	CrGR1
Data collection	
Unit cell	a=77.5 Å, b=99.7 Å, c=107.8 Å, α= 90°, β=90°, γ=90°
Space group	P 21 21 21
Resolution range* (Å)	73.215-1.713 Å (1.743-1.713 Å)
Unique reflections	59484
Completeness* (%)	100.0 (99.2)
R_{merge}*	13.0 (108.9)
I/σ	13.9 (2.2)
Multiplicity	12.7 (13.3)
Refinement	
Resolution range* (Å)	73.21 - 1.71 (1.758-1.713)
Reflection used	85749 (6240)
R/R_{free}* %	19.49/24.4
rmsd from ideality (Å, °)	0.021, 2.209
Non-hydrogen atoms	7830
Protein atoms	7342
Water molecules	438
B value (Å^2)	
Mean (chain A and B)	28.081
Solvent molecules	30.00
Ramachandran plot (%)	
Most favored	94.88
Allowed	3.38
Outliers	1.74

4.4.2 Choice of the best model and resolution

Once the full set of data was collected, main output parameters were evaluated to find the most representative data in terms of statistics. In that case overall resolution estimate, completeness, I/σ, and shell multiplicity are crucial. The choice here would mean picking the best data for further processing and again the data collected for III/C6/33B were found to be of an excellent quality compared to others. Then *mtz* file containing the amplitudes of structure factors and errors was extracted and taken for structure determination procedures. From the experimental data, it is possible to guess the asymmetric unit content by calculating the Mattews coefficient and the solvent content. For GR1 the value of the Mattews coefficient equal to 2.32 indicates that a whole dimer is contained in the asymmetric with a solvent content of 47%.

As mentioned in the previous chapter a lot of information is automatically processed, including the structure factor F_{hkl}. During an XRD experiment, the information about the phase of a diffracted beam is lost, meaning there is no way to assign the origin the wave itself.

Molecular replacement was chosen as the first attempt for the phase problem solution. This approach is possible when homologous protein is available in PDB (thus with at least 30-35% identity). Here this is the case since quite few protein structures are reported for GR. As a first shot the GR with the best identity (~40%) was taken for structure determination - *Sinorhizobium meliloti* (PDB ID: 4DNA, Uniprot: Q92PC0). One protein chain deprived of ligand and water molecules was used. The calculations gave two solutions, corresponding to two protein chains, thus a whole dimer, in the asymmetric unit as indicated by the Mattews coefficient (unit cell packing shown in Figure 18). At the first sight the model seemed to be working to some extent, but unfortunately, after few

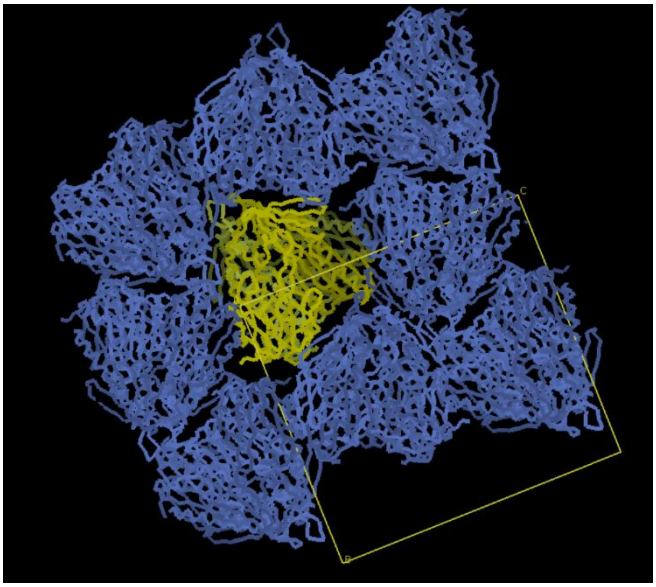


Figure 18. Unit cell packing for GR2. The neighboring proteins did not show clashes, hence molecular replacement worked properly.

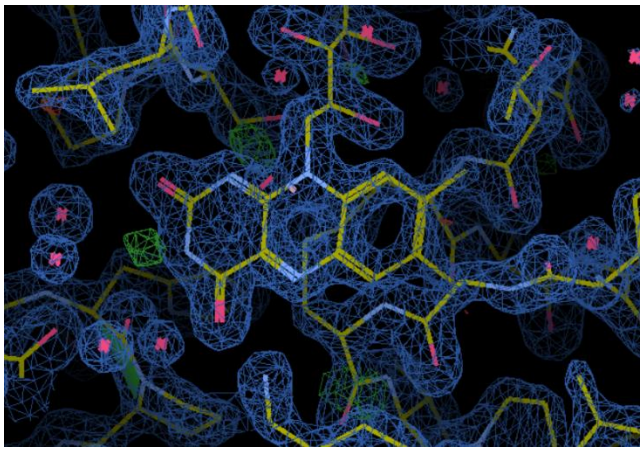


Figure 19. $2F_0$-F_c electron density map contoured at 1σ around FAD region. Visualized with Coot.

attempts, there were regions with difficult-to-identify densities and the refinement calculations stacked. Another problem that appeared was the chain length. The difference of thirty amino acids (model had 463 residues while GR1 has 493) made it difficult in the model building at the C-terminal end.

Thus, the structure of GR from *Saccharomyces cerevisiae* (PDB ID: 2HQM, Uniprot: P41921) deprived of ligand and water molecules, was taken as a model for the second molecular replacement attempt. Its identity with GR1 was 32% but the chain length was 473 residues. Thus it was a convenient solution in terms of fitting amino-acids into the density, and no unassigned electron densities were observed. Resolution better than 2 Å simplifies the work in terms of the model building because all residues can be distinguished as in Figure 19 where the flavin ring of a FAD nucleotide exhibit clear, assignable density. Final refinement statistics are reported in Table 5.

The resolution determines the quality of the electron density, final structure and the resolution itself is purely dependent on the quality of the crystal.

4.5 FAD binding pocket and comparison with pro- and eukaryotic organisms

Final resolution of the *Chlamydomonas reinhardtii* GR was 1.7 Å allowing for reliable structural insights. GR is a homodimer employing FAD and NADPH to efficiently convert one molar equivalent of GSSG into two equivalents of GSH. The schematic mechanism is based on proton donation/acceptance as presented in introduction chapter. The enzymatic mechanism of GR is of ping-pong type and is connected to reductive and oxidative half reaction.

Redox chemistry involved in GRs enzymatic activity has certain well-conserved residues, such as histidines and cysteines. Understanding flavoenzymes — enzymes using FAD in catalysis — is particularly complex because of flavin's involvement in a wide variety of chemical reactions and its possible redox and protonation states, each of which has

unique properties and can be stabilized or destabilized by the protein environment. [32] A comparative study of three-dimensional structures and sequences stands for a good indication of evolutionary pathways. Even though GR is an extensively studied enzyme there was no single plant-origin structure available and this is probably the first attempt in comparing protein-ligand interactions for GR from five kingdoms representatives, being:

I. Plantae - *Chlamydomonas reinhardtii*
II. Animalia - *Homo Sapiens*
III. Fungi - *Saccharomyces cerevisiae*
IV. Protista - *Plasmodium falciparum*
V. Bacteria - *Rhizobium meliloti*

Where in yellow are highlighted eukaryotic organisms and in blue prokaryotic. This group of organisms could be considered complete family around ten years ago. Pace *et al*. (2006) in his paper *Time for change* (Nature) [34] classified Archaea as a distinct kingdom, thus we should include one more prokaryotic kingdom. Yet it can be considered a representative set of structures. In terms of interactions of the flavin with the neighboring residues the most interesting ones include tyrosine, cysteine disulfide bond, and histidine since the first is responsible for accommodation of NADPH (as stated in the introduction - tyrosine is a ‘gate-keeper’), cysteines are here important in terms of histidine protonation/deprotonation steps that further facilitate conversion of GSSG into GSH.

The main difficulty in structural studies of enzymes employing many protonation/deprotonation steps is the resolution. Even structures at 0.95 Å do not provide for example protonated state of the histidine. [33] The proceeding subchapters present graphically interaction of FAD with its environment with highlighted distances.

4.5.1 Plantae - *Chlamydomonas reinhardtii*

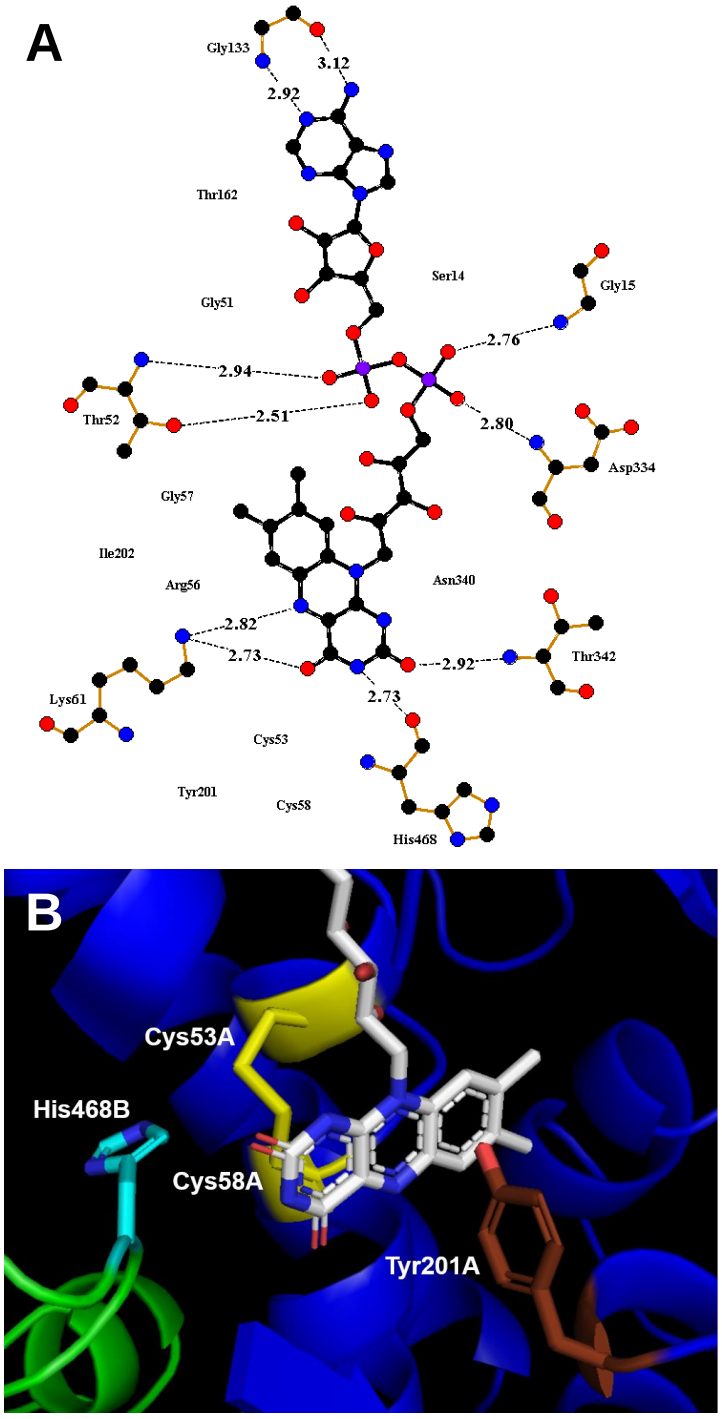


Figure 20. Image A: ligand interaction with neighboring residues and hydrogen bonds distances. Image B: three dimensional view on FAD binding pocket with highlighted catalytic cysteines (yellow), histidine (cyan), and tyrosine (brown).

The resolved structure of plant origin GR shows well conserved structural features. The FAD pocket shows a Rossman fold characteristic of proteins that bind nucleotide as NAD^+, $NADP^+$ or FAD. In GR1 is formed by parallel β sheets and 5 α helices. The first β-strand is followed by a tight loop that is connected to the N-terminal of the helix. The two highly conserved Gly residues in the sequence are located in this turn to allow the sharp bending of the chain. At the end of the helix there is a wider turn that is followed by the second beta strand that runs parallel to the first strand. Here the Rossman fold holding FAD molecule is stabilized by same hydrophobic side chains. Also, those involved in catalytic activity seems to have very similar spatial orientation compared to other GRs as can be seen in Figure 20.

4.5.2 Animalia – *Homo Sapiens*

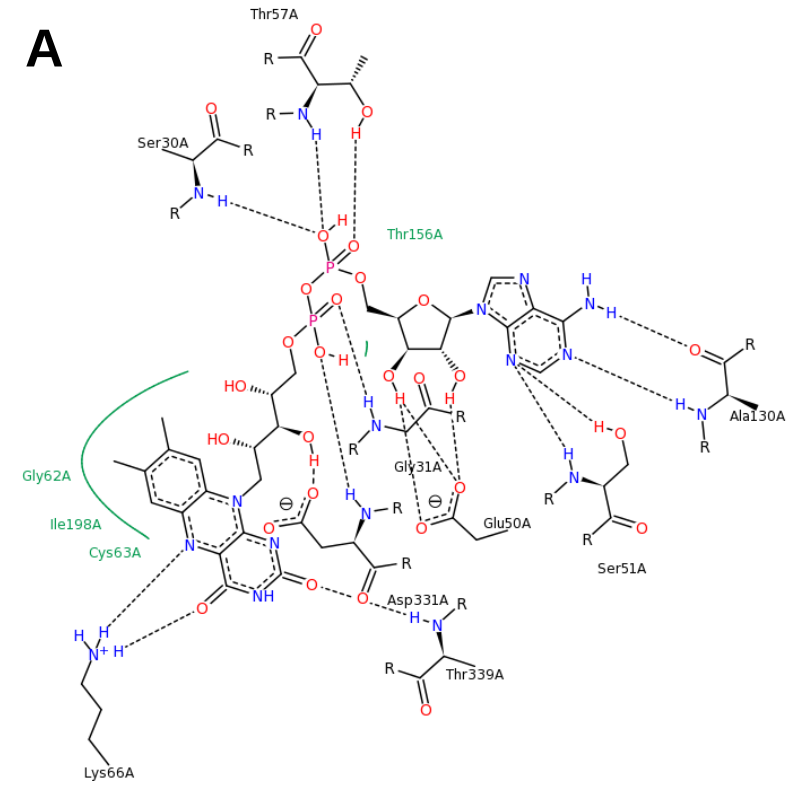
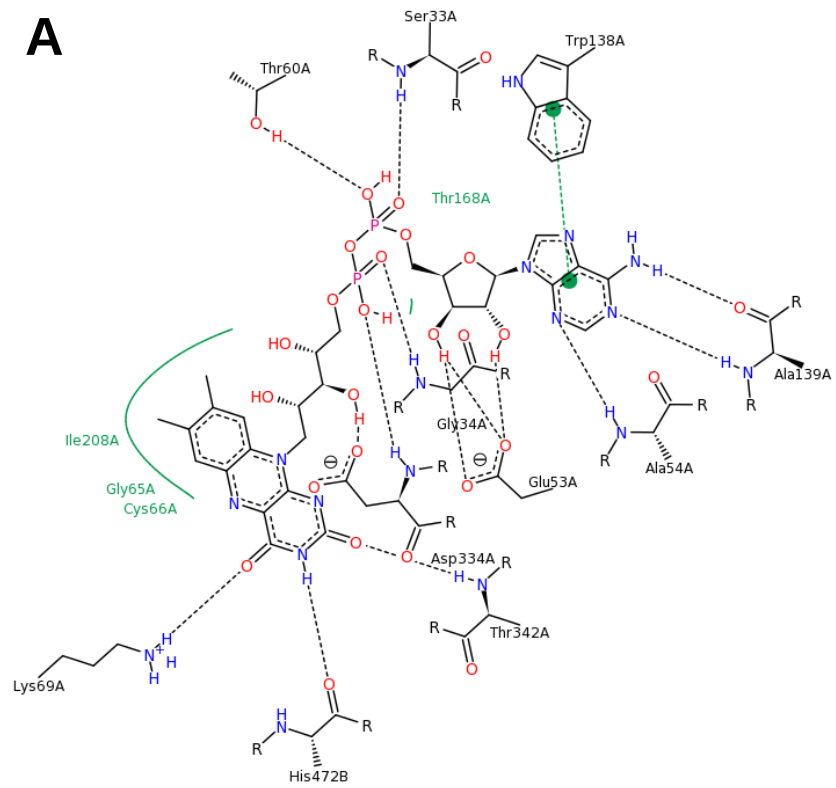


Figure 21. Image A: ligand interaction with neighboring residues and hydrogen bonds distances. Image B: three dimensional view on FAD binding pocket with highlighted catalytic cysteines (yellow), histidine (cyan), and tyrosine (brown). Accession code: 3DK9.

Animal origin GR was actually the first GR resolved. It gives 43 hits in PDB (including transferases, nitroso-GRs, carbonyl-GRs and hydroquinones) database. In this study the best resolution GR was taken (0.95 Å). Figure 21 presents two-dimensional view on FAD binding pocket with hydrophobic residues and the three dimensional representation.

4.5.3 Fungi - *Saccharomyces cerevisiae*



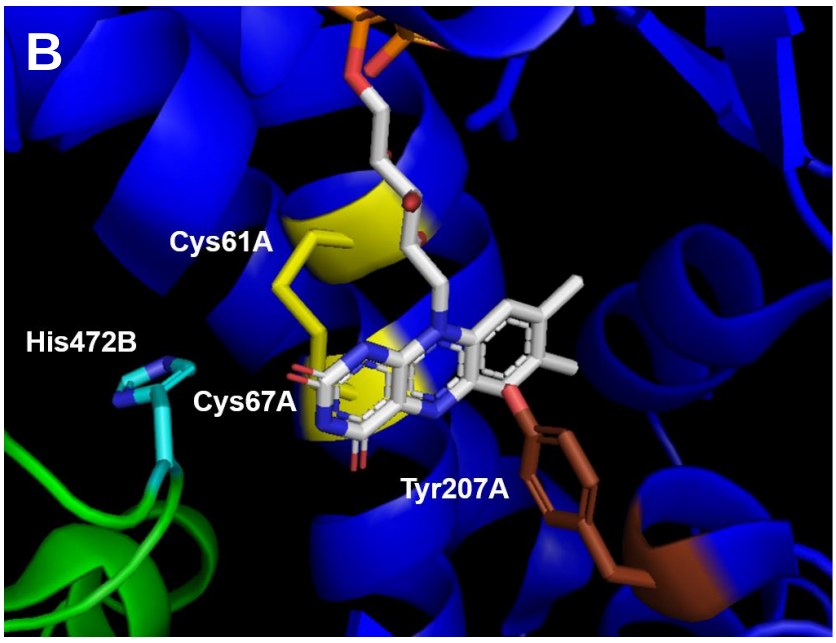
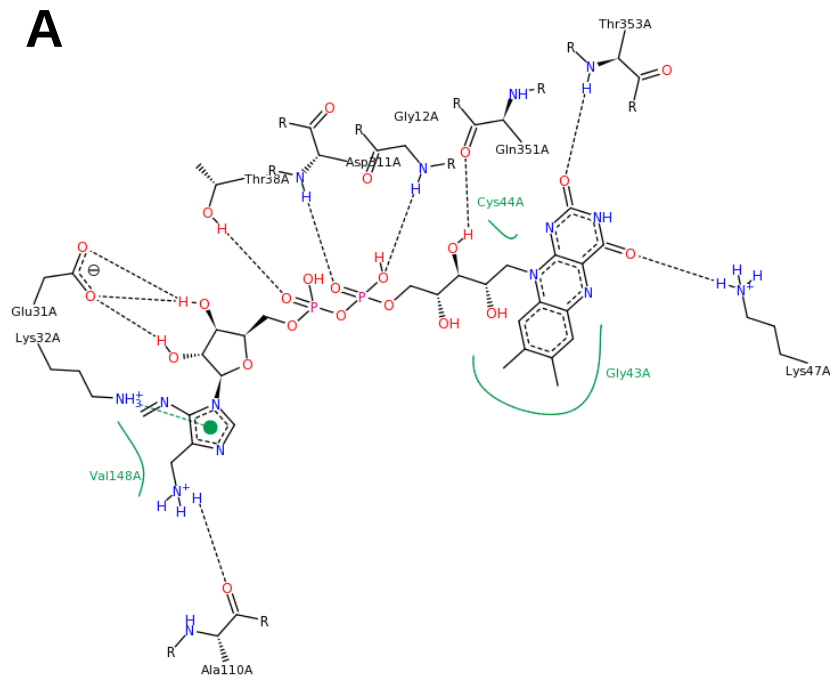
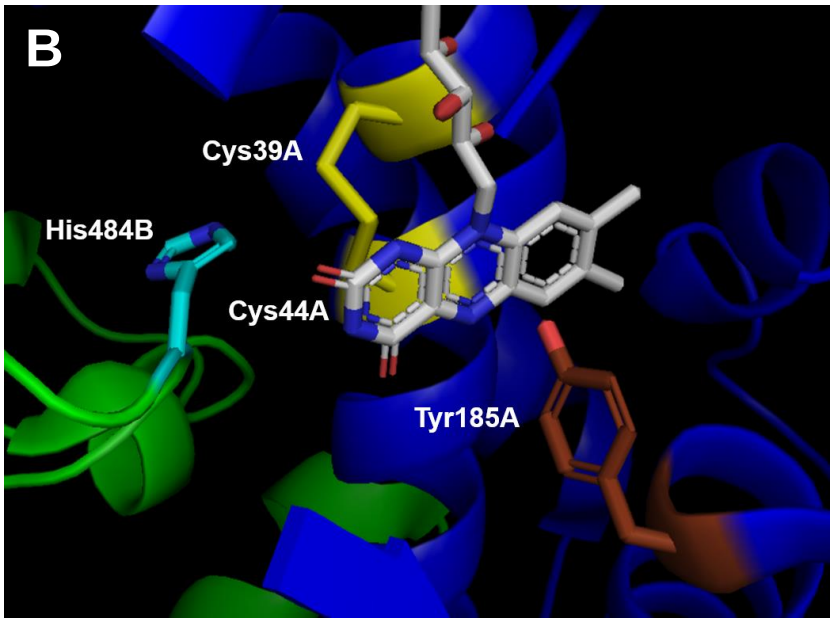


Figure 22. Image A: ligand interaction with neighboring residues and hydrogen bonds distances. Image B: three dimensional view on FAD binding pocket with highlighted catalytic cysteines (yellow), histidine (cyan), and tyrosine (brown). Accession code: 2HQM.

Fungi GR is one of the few resolved structures accessible on PDB and was resolved at 2.4 Å. The first yeast structure was reported in 2007 under PDB ID: 2HQM. Figure 22 presents neighboring residues interaction in two- and three dimensions respectively.

4.5.4 Protista - *Plasmodium falciparum*

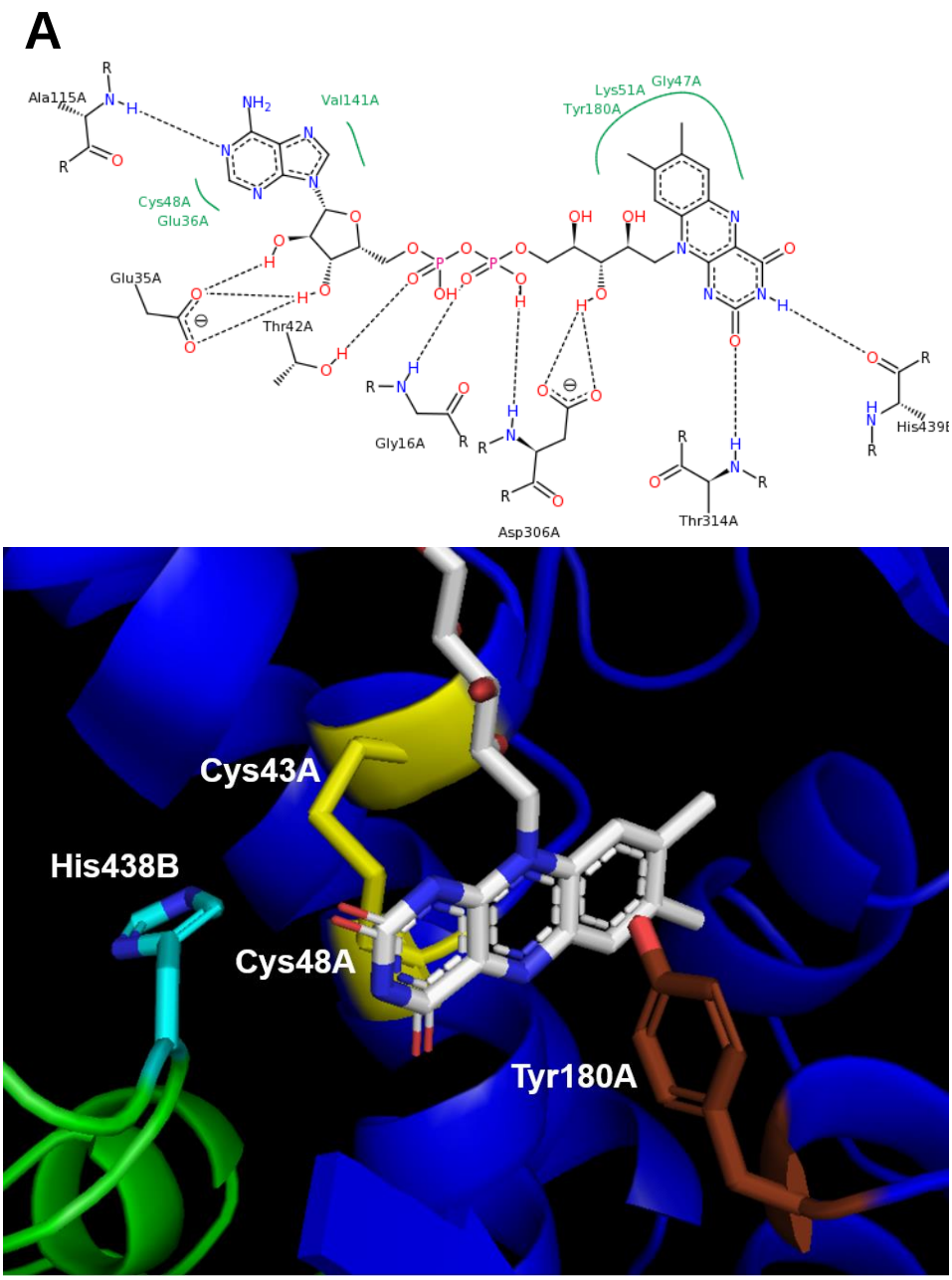




Protista GR representative shown here in Figure 23 and comes from structure resolved at 2.6 Å in 2003 with accession code 1ONF on PDB.

Figure 23. Image A: ligand interaction with neighboring residues and hydrogen bonds distances. Image B: three dimensional view on FAD binding pocket with highlighted catalytic cysteines (yellow), histidine (cyan), and tyrosine (brown). Accession code: 1ONF.

4.5.5 Bacteria - *Rhizobium meliloti*



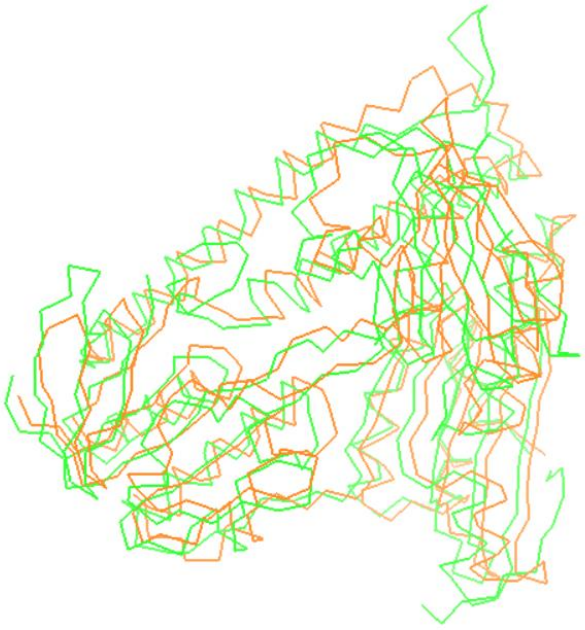
Bacterial GR is extensively studied enzyme and most popular species would be *E. Coli* but highest identity with plant structure was observed for *R. Meliloti*. Again without a surprise many residues maintain similar orientation around FAD nucleotide and are presented on Figure 24. This structure was obtained at 2.8 Å resolution and submitted to PDB in 2012 with an accession code 4DNA.

Figure 24. Image A: ligand interaction with neighboring residues and hydrogen bonds distances. Image B: three dimensional view on FAD binding pocket with highlighted catalytic cysteines (yellow), histidine (cyan), and tyrosine (brown). Accession code: 4DNA.

4.5.6 Structural superimposition

The conclusion withdrawn from the catalytic site comparison among various origin GRs is that they are very well conserved. Thus the question arise, where the difference in identity come from. In order to compare the whole structure of different GRs, a superimposition of Cα atoms was performed. Figure 25 presents the superimposition of *Chlamydomonas reinhardtii* GR1 with human GR and bacterial GR.

Red: *C. reinhardtii* GR
Blue: *H. sapiens* GR
r.m.s.d. 1.22 Å



Green: *C. reinhardtii* GR
Orange: *E. coli* GR
r.m.s.d. 1.39 Å

Figure 25. Structural alignment of single GR subunits. (A) Plant and human GR. (B) Plant and bacterial GR.

Hence while looking at root mean square deviations (r.m.s.d.) and conservation of the secondary structure elements it is seen that they are spatially related, while in plant and bacterial subunit differences are more pronounced. Loops are the regions that contribute substantially to the identity difference, as being regulatory elements in many enzymes. Many loops are exposed to the outer environment in a cell and/or the interface in dimers. For GRs probably both aspects contribute notably, what was observed at the model building step. The electron densities in such regions are of lower quality in majority of the cases, meaning they are not rigid fragments (even in the crystal). This dynamic behavior along with chain length differences account for visible difference in both structure and protein identity.

4.6 Conclusions

Despite the extensive enzymatic and structural studies of GR catalysis, uncertainties remain that more detailed pictures of catalysis might resolve. For example, in the human GR native structure determined at 1.54 Å resolution, marginally reliable deviations in peptide planarity of the active-site disulfide loop were suggested to indicate strain that would favor disulfide reduction, as reported by Berkholz et al. (2008). Slight deviations in flavin geometry were suggested as possible evidence as to how the protein modulates flavin reactivity. Also, the more accurate determination of the geometry of the nicotinamide-flavin approach provided by atomic-resolution analysis would help in better understanding of the hydride transfer step. Finally, the atomic-resolution analysis could yield additional insight by providing a direct visualization of the protonation states of the active site Cys and His residues at various stages of catalysis.

High-resolution structure determination is also essential for precise mechanism assessment because as stated in the report by Weik *et al.* (2000) synchrotron radiation can cause specific structural changes such as the cleavage of disulfide bridges and reduction of active site metallocenters. This has one advantage of providing a view of catalytically relevant reduced enzyme forms that can allow insights into enzyme mechanism. But on the other hand, some groups state that radiation-induced reduction differs from those naturally reduced enzymes thus taking away valuable information in

terms of the geometry of catalytic sites. One of the solutions would be an attempt with NMR studies or hydrogen/deuterium exchange experiment.

For plant glycolytic GR most of the Rossman fold features are conserved even if the sequence alignment does not exceed 41% with any other GR available. The FAD-binding pocket and its catalytic residues, being His468B, Cys53A and Cys58A are shown in Figure 20.

Very interesting observation is made while studying the mechanisms and ping-pong type behavior of NAPDH. Literature, in this field, is not very consistent, since some reviews state that those are very clear, while the other groups are very uncertain about structural features. The proton donor (NADPH) enters and leaves its site many times during catalytic cycle as described in *Introduction* (redox half-reactions). On PDB only one group (at least to the author's knowledge) reported NADPH state of GR of human origin (PDB ID: 3DJG). This opens up many possibilities in the determination of catalytically active steps structural changes. Reduction of GSSG disulfide involves few protonation and deprotonation steps, a formation of charge transfer complex, proton donor exchange, and all in assistance of neighboring residues affecting the kinetics of the S_N2 reaction.

Last but not least concern is more general thought on how many molecules of GSSG are bound at once to the GR? One would say, two since GR is a homodimer and hence has two catalytic sites. Considering complexity of this process it is not becoming very clear, whether one GSSG is bound at a time and then after release of two GSH molecules, second site becomes available for the new reaction? This and few others become truly interesting directions in the flavoenzymes research.

4.6.1 Perspectives

Compared to available structures of GR the resolution of 1.7 Å can be considered a good achievement and also motivation for the continuation of this project. There are certain possibilities of expanding this study by determining the structure with $NADP^+$ and NAPDH state of the enzyme. Structure with reducing NAPDH would also grant some dynamic insights about Tyr201 and its behavior upon nucleotide binding.

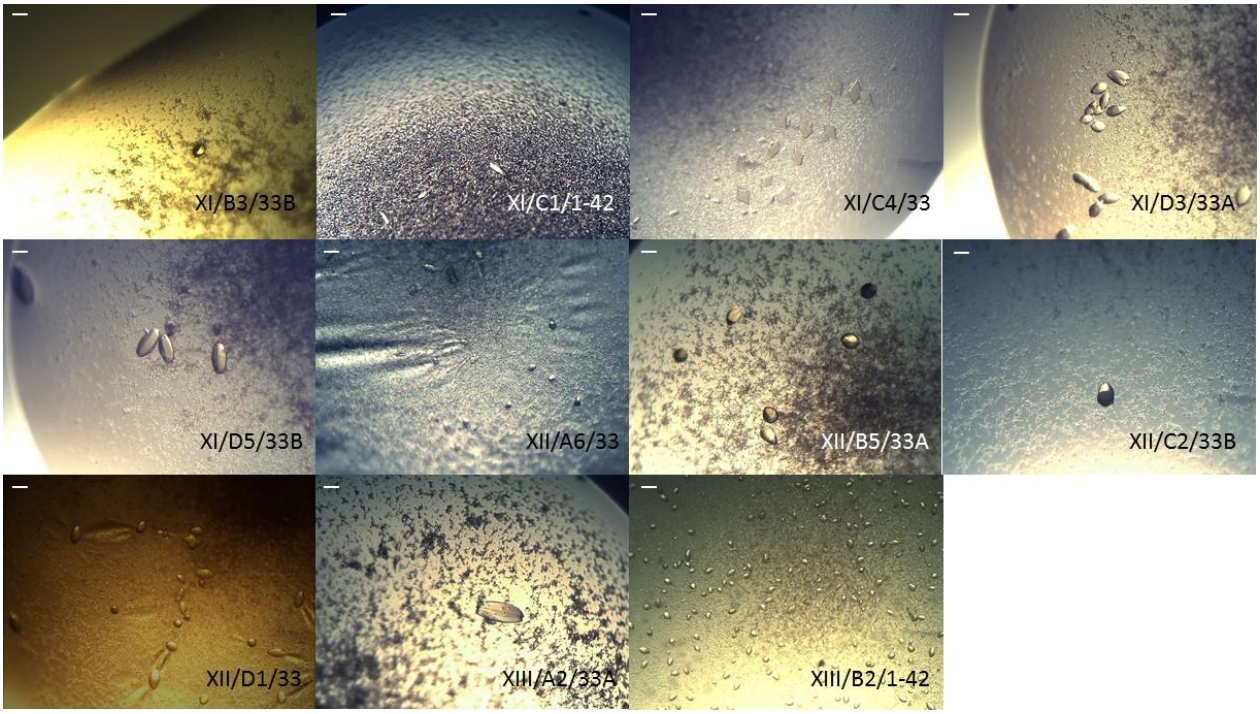


Figure 26. GR1 crystals. From the top left corner: first two images are GR1 with $NADP^+$, next three images are GR1 reduced with NAPDH, next four images are GR1 with TCEP and remaining two are GR1 with NAPDH and TCEP. White scale bar in every corner has approximate length of 70 µm.

At this stage, having steady protocols for GR1 crystallization, it is possible to start crystallization process with other cofactors. Figure 26 presents successful crystallization conditions of plant GR1 with $NADP^+$, NADPH and TCEP (a phosphine commonly used for disulfides reduction). Those are likely going to be taken for XRD analysis at synchrotron radiation source. The morphology of the crystals, as being seed-like shaped, will be probably little improved after minor optimizations. Each of those has to be crystallized separately to avoid catalytic activity. Another idea for the project is to co-crystallize GR with GSH and GSSG without NADPH to identify the affinity of each subunit for glutathione binding. Difficult to guess the how successful this would be, hence complementary techniques will need to be employed (such as molecular docking).

PROJECT2

Chapter 5 - Introduction

5.1 Importance of BODIPY fluorescent probes

Live cell imaging had become fascinating research area in the light of advanced microscopic techniques recently developed. Fluorescence microscopy is an integral part of cell biology and the number of tools for tagging and imaging is constantly increasing. Nevertheless, in cell imaging there are few challenges difficult to overcome, that is minimizing photodamage of the emitting probe, keeping fair signal-to-noise ratio, maintaining physiological conditions in the living cell to study dynamics in a stable environment and notably choice of the fluorescent probe. [35]

Fluorescent probes as being available in hundreds, if not thousands, still have properties that generate a number of issues when it comes to imaging of biological matter. Recently developed boron-dipyrromethene (BODIPY) family of fluorescent dyes have excellent photophysical properties and gained recognition in chemistry, molecular biology, and physics. Ease of tuning and hence shifting the emission spectra is one of the numerous advantages of this class of compounds.

5.1.1 Fluorescent probes in cell imaging

Modern cell biology aims at studying dynamic processes within the living cell, such as signaling, cellular excitability (i.e. receptors), growth and/or metabolic processes. This requires set of tools that can effectively provide a valuable information. Cell dynamics is not a trivial target, it requires temporal resolution in both photometric and image data. Experimentally there are two aspects to be considered, that is sample preparation and design of the fluorescent probe.

Ultimately the objective is to have a molecule that targets a region of a cell and gives visible emission under irradiation with light as a stimulus. In very few cases intrinsic probes

can be informative since their location is rather specific and may give a solid signal, for instance, NADH, flavins, and porphyrins – endogenous fluorophores.

In the design of the probe there are few criteria to be fulfilled:

- Delivery. The way fluorescent probe is introduced and distributed in the living system cannot affect the physiology and structure of the cell. The chemical character play here crucial role. If the probe has a lipophilic character it can enter the cell only by the addition of small-volume water-miscible stock solution. Cell population analysis with flow cytometry or spectrofluorometry requires bulk loading procedures, such as membrane-permeant acetate, acetoxymethyl ester derivatives (Tsien, 1981), acid loading (Bush and Jones, 1990), hypo-osmotic shock (Nolan *et al.*, 1988) or electroporation (Bright *et al*.,1996). One interesting development in the field is the application of cationic liposomes to load the F-actin probe BODIPY FL phallacidin into live neuronal cells (Barber *et al.,* 1996).
- Targeting. The probe should come into interaction with a biological sample in a selective manner. Common visualization of this aspect is to think about the probe as a molecule involving two functional compartments – a fluorophore and a targeting group. Both are equally important, nevertheless quality of the fluorophore will determine image resolution – this is why ease of functionalization is of high importance.
- Detectability. Photophysical properties of the probe must be consistent with the instrumentation that is applied for the detection. The signal must be sufficient at concentration levels that correspond to the natural abundance in the cell. Detectability is closely related to a few spectroscopic characteristics of the probe being: extinction coefficient and fluorescence quantum yield, photobleaching, excitation wavelength and, Stokes shift.
- Fluorescence response. This describes how the interacting probe is going to be detected (e.g. imaging spatial distribution of fluorescence or spectroscopic response such as spectral shift). Intracellular environment usually affects fluorescence sensitivity (even though there cases where its lack is advantageous). Globally this can be used to generate a change of fluorescence intensity in response to the interaction of the probe with its target. [36]

Small molecules as fluorescent probes possess few assets over fluorescent proteins (FPs) - most importantly their size, ease of functionalization and Near Infra-Red (NIR) fluorescence. [37] Widely used fluorophores include coumarin, fluorescein, BODIPY, rhodamine and cyanine dyes (Figure 27) and cover a wide range of fluorescence depending on π-conjugation and effect of a moiety attached.

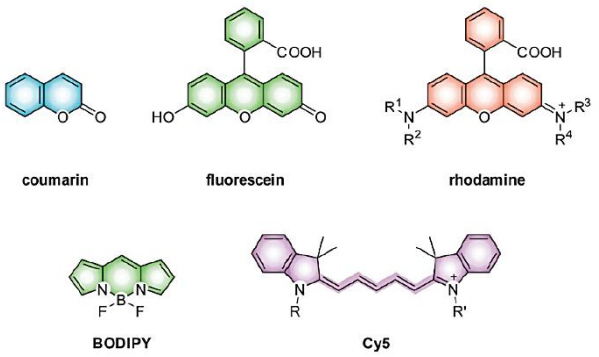


Figure 27. Examples of small molecules widely applied in cellular imaging as a fluorescent probes.

5.1.2 Structure and physicochemical properties of BODIPY dyes

4,4-Difluoro-4-bora-3a,4a-diaza-s-indacene abbreviated as BODIPY is a compound with interesting electronic properties and a large variety of derivatives. The chromophore structure and structure investigated in this work are presented in Figure 28. The advantages are quite numerous and application range includes biomolecular labels, drug delivery agents, fluorescent switches, electroluminescent films, dye lasers, light-harvesters, and sensitizers for solar cells. [38]

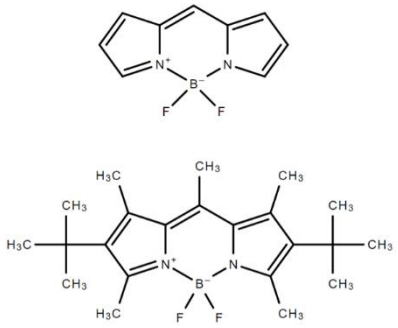


Figure 28. Top: chemical structure of the chromophore of BODIPY. Bottom: BODIPY utilized in this work (also referred as pyrromethene597).

Significant in biochemical/biological research property is a strong UV absorption and a relatively sharp emission spectrum with a quantum yield close to 1. One of the first BODIPY structures was reported in 1968 by Treibs *et al*. [39] but the first non-functionalized core was synthesized in 2009 and published independently by three groups. [40][41][42] These dyes are insensitive to the polarity and pH of the environment (as a general trend, with certain exceptions reserved for some derivatives). The BODIPY chromophore is relatively non-polar and electrically neutral molecule with planar geometry. The emission profile is tuned easily upon binding to proteins or DNA. Another important advantage is a small Stokes shift, making them, as mentioned before, a good match for imaging. Focusing specifically on cellular biology

there are drawbacks of those. Most of the BODIPY's emit at wavelengths lower than 600 nm and only handful of water-soluble dyes are commercially available. The potential of those is yet unrealized in living cell imaging and further developments are required. Their chemistry was extensively explored and synthetic routes are described in a review by Loudet *et al.* (2007) [47] explicitly for molecular biology applications. Many states that high photostability of pyrromethenes makes them superior to rhodamine dyes, texas red or fluorescein. [43][44]

Some BODIPY derivatives exhibit specific affinity toward some amino acids. [51] The selectivity of the probe towards certain amino acids was assessed by using optical spectroscopies (here absorption and fluorescence). Spectral changes upon titration (from fluorescence emission intensity evolution) revealed that BODIPY derivatives have a strong affinity for cysteine and homocysteine.

5.2 Synthesis and functionalization of BODIPY dyes

5.2.1 From pyrroles and acid chlorides or anhydrides

A common approach to get functionalized core of BODIPY is to hit meso position first via a condensation reaction of acyl chlorides with pyrroles. Such reaction involves unstable intermediate of dipyrromethene hydrochloride salt formation which makes a little easier to handle and purify due to increased C-substitution. Reaction scheme is presented in Figure 29A. Acid chlorides are not unique here and can be substituted with other carboxylic acids derivatives. Another approach includes glutaric anhydride as a starting reagent and has an interesting feature of free carboxylic acid at meso position, which can be further used to attach the dye to a target molecule. The scheme presented in Figure 29B.

5.2.2 From pyrroles and aldehydes

The similar synthetic approach combines oxidation steps and aromatic aldehydes as reagents. Oxidation reactions can possess some experimental inconveniences, such as byproduct removal. Figure 29C shows such reaction with use of p-chloranil as an oxidant.

If one aims at obtaining α,β-unsubstituted BODIPY, it can be done starting from aldehydes and is believed to be an elegant approach requiring mild conditions. It is enough to dissolve aldehyde in excess of pyrrole at 25 °C and dipyrromethene intermediate is obtained (and furtherly isolated). The final product is formed during oxidation with 2,3-Dichloro-5,6-Dicyanobenzoquinone (DDQ) and complexation with boron. The scheme is shown in Figure 29D.

The broad scale of possibilities would offer use of halogenated benzaldehydes as for compound in reaction 3. This contributed to the development of a long chain acid substituents and in fact were successfully applied to study dynamic effects in membranes (structures on Figure 29E). Moreover, ortho-substitution on the meso-aromatic groups restrict moiety rotation and contribute to higher quantum yields.[45] For this approach number of possibilities are reported in the literature and their choice depends purely on purpose. Shortly, condensation reactions of pyrroles with either acid chloride or benzaldehyde derivatives are considered convenient and direct methods for symmetrical BODIPY preparation.

5.2.3 From ketopyrroles

Last synthetic route outlined here is devoted for unsymmetrically substituted BODIPY. This includes (generally) preparation of ketopyrrole intermediate followed by a Lewis acid mediated condensation with another pyrrole fragment. Magnesium anions of pyrroles upon reaction with an activated carboxylic acid derivative give 2-ketopyrroles.[46] Then it is followed by condensation of ketopyrroles with another pyrrole to result in BODIPY structure. Even though it gives mainly symmetrical product this method is also applied for unsymmetrically substituted ones (Figure 29F).

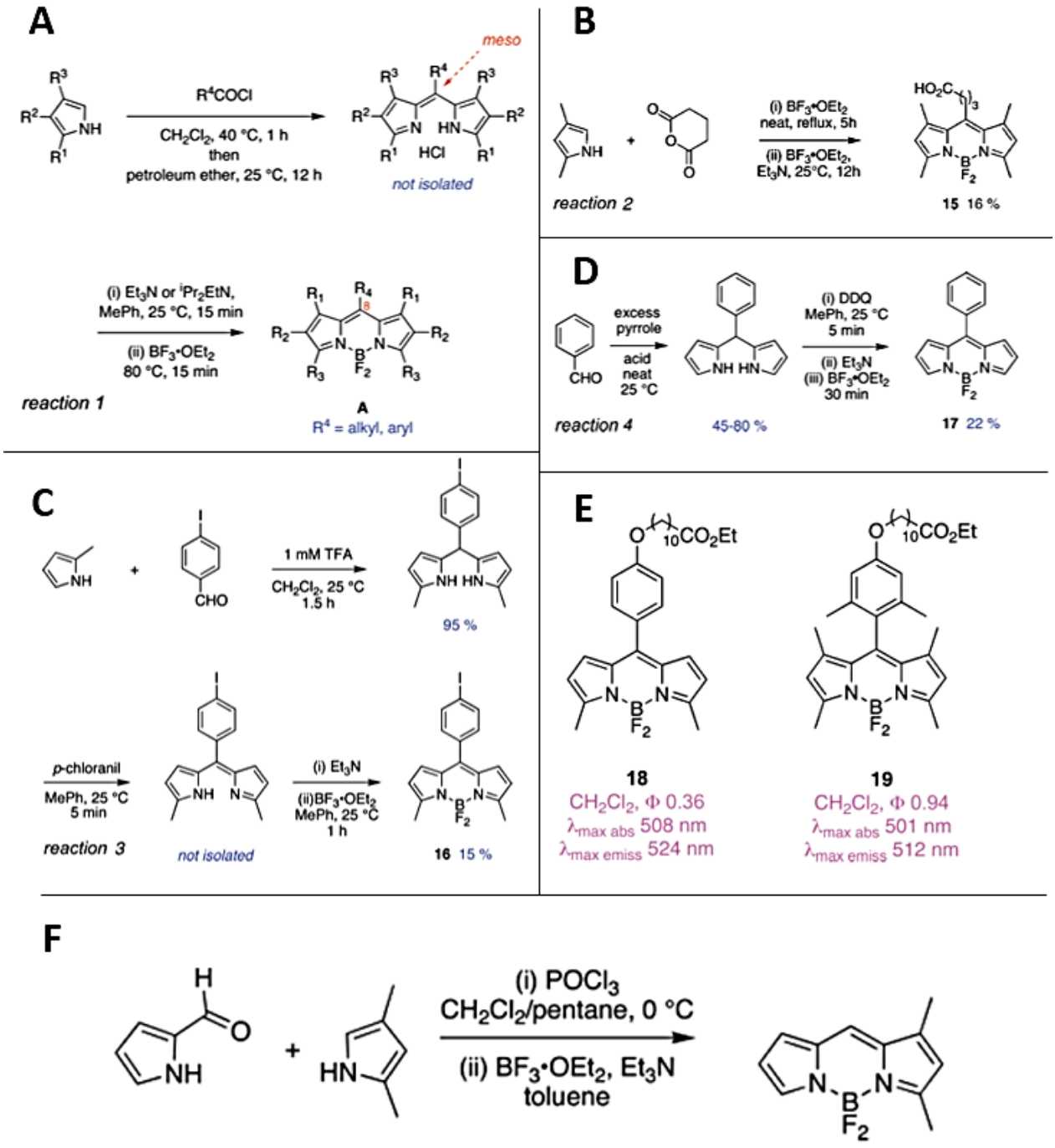


Figure 29. Synthetic routes for BODIPY functionalized chromophore. Description of reactions in the text in chapter 5.2. Reaction schemes taken from [47].

5.3 Bovine Serum Albumin as a model protein

One of the issues when dealing with BODIPY dyes is the solubility. Hydrophobic variants of BODIPY are easily available and relatively low-priced compared to functionalized ones. Still, there is another way round to overcome this problem. Incorporation of the probe into a protein would own a great advantage of water solubility and hence delivery problems with no need for harsh chemical conditions and provide an instant idea about targeting since the protein would be localized in protein-specific organelles. In PDB there is only one structure of this kind, under accession code 2ZK6. It is a structure of a human peroxisome receptor with functionalized BODIPY derivative. Even though that study concerns a receptor with multiple affinities, it has a highly helical structure, thus BSA might be an interesting target too.

Serum albumins (also known as blood albumins) are well-known proteins found in vertebrate blood. Mammalian serum albumins are believed to be chemically similar. They are synthesized in the liver and spread into blood plasma (in fact it is most abundant blood protein in mammals). Albumin is synthesized in the liver as preproalbumin which has an N-terminal peptide that is removed before the nascent protein is released from the rough endoplasmic reticulum. The product, proalbumin, is in turn cleaved in the Golgi vesicles to produce the secreted albumin. [48] There they are responsible for maintaining the oncotic pressure that is essential for delivery of body fluids between blood vessels and body tissues. It also acts as a transporter by non-specifically binding hydrophobic steroids, thyroid hormones, hemin and fatty acids.

Serum albumins show several advantages such as being globular proteins, stable in terms of folding, water-soluble, and structurally well characterized. For all these reasons they are very often used as model proteins. From more practical view Bovine Serum Albumin (BSA) is a commercially available material. In biochemistry, immunology and biotechnology BSA is commonly used in Enzyme-Linked Immunosorbent Assay (ELISA), immunocytochemistry, immunoblots and in a stabilization of restriction enzymes. Due to lack of interference with many biological reactions, it is used as a universal blocking agent.

The molecular weight of BSA lays around 66 kDa with a α-helical structure (maintaining static shape for blood pressure regulation) – Figure 30. Within serum albumin structure there are eleven hydrophobic pockets for hemin and fatty acids. In total there are seven sites for long and medium sized fatty acids (Ashbrook *et al.*, 1972, 1975), four sites for short fatty acids (Bhattacharya *et al*., 2000), two sites for drugs (Sudlow *et al*., 1975, 1976), one site for bilirubin (Zunszain *et al*., 2008) and, five sites with affinity for thyroxine (Petitpas *et al*., 2003). All albumins are cysteine-rich proteins (thus are interesting targets for BODIPY in terms of the cysteine affinity) and sequence alignment for BSA, HSA, ESA, and LSA had shown that their similarity level is very high. First three-dimensional structure of BSA was deposited in 2012 at resolution 2.47 Å (PDB ID: 4F5S).

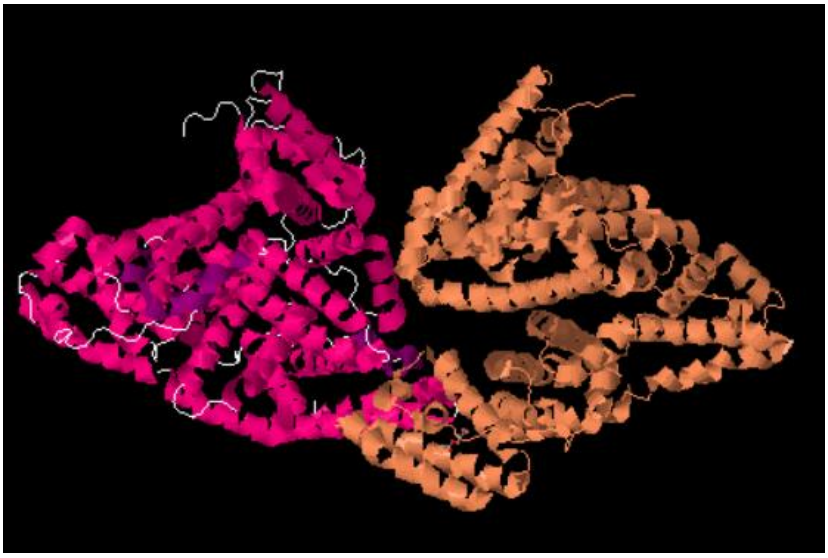


Figure 30. Crystal structure of BSA subunits.

Chapter 6 - Methods and experimental fine points

6.1 Chemicals and sample preparation

Bovine serum albumin (BSA) in a form of lyophilized powder was purchased from Sigma-Aldrich (St. Louis, MO, USA), and pyrromethene597 was purchased from Exciton (Dayton, OH, USA).

Adduct preparation of pyrromethene597 and BSA was varying a little from case to case, but as a general trend ratio of protein:probe was between 10 and 20 fold dye excess with different stirring methods (magnetic stirrer, rocking platform, bath and microtip sonication):

Method 1: For an absorption experiments the protocol incorporate protein:dye at ratio ~1:20 (0.25 mM for BSA and 5.5 mM for pyrromethene597 in water) and left under magnetic stirring for 48 hours. Then sample was centrifuged (10.000 rpm, 4 °C for 10 minutes) and taken for DLS. Protein reference solution had concentration of 0.25 mM (in water).

Method 2: In the second protocol applied in both absorption and fluorescence experiments protein:dye ratio was kept around ~1:10 (0.5mM for BSA, 5mM for pyrromethene597 in water). One fraction of the sample was taken for sequential sonication (5x10 minutes cycles of microtip sonication - in total 50 minutes) and remaining fraction for magnetic stirring (24 hours). Then samples were centrifuged (10.000 rpm, 4 °C for 10 minutes) and taken for DLS measurement.

6.2 Absorption and fluorescence spectroscopies

Absorption and fluorescence spectroscopies belong to the family of optical techniques aiming at the determination of various sample properties, such as concentration and photo-physical properties. Both are based on photon interaction with a dye/protein molecule, being here absorption and subsequent emission.

Overview of the processes in a simplified way is presented in Figure 31. The fluorophore (dye molecule or aromatic residues of the protein) can be in two different states with specific energy gap ΔE. Absorption occurs when an incident photon of matching wavelength (or energy) hits the molecule in the ground state and brings it to the excited state so that the photon is absorbed in the process. The condition of the photon to be absorbed must fulfill the following energy relation:

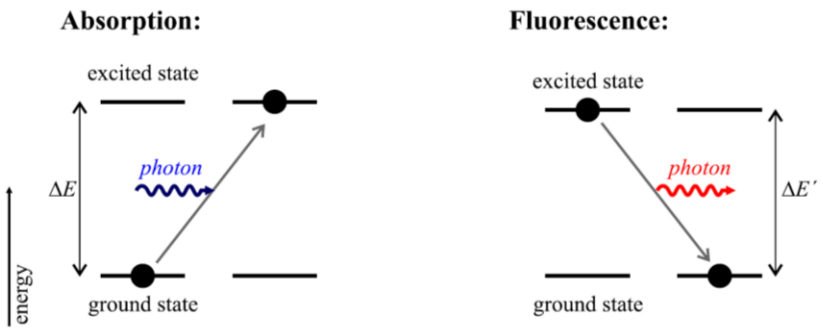


Figure 31. Absorption of an incident photon and emission of a fluorescence photon in a simplified fluorophore electronic state system.

$$E_{photon} = \Delta E$$

Fluoresce is a process altering absorption. If the excited molecule returns to its ground state, the energy $\Delta E'$ has to be deposited in some way. One of the possibilities is a radiative relaxation, or simply emission of a fluorescence photon carrying the energy of the gap $\Delta E'$. The way the energy is dissipated from the excited state depend on the dye character, and in the case of the fluorescent probe, it is a dominant path.

6.2.1 Absorption theory and experimental conditions

Most elementary law describing the effect of an absorption is Beer-Lambert law in the case where light passes through the material (sample). The decrease in transmitted light (absorbance) is connected with the material properties. The scheme of an absorption measurement is depicted in Figure 32. This law may be rewritten in terms of an absorption A that express relative decrease of intensity in a logarithmic scale:

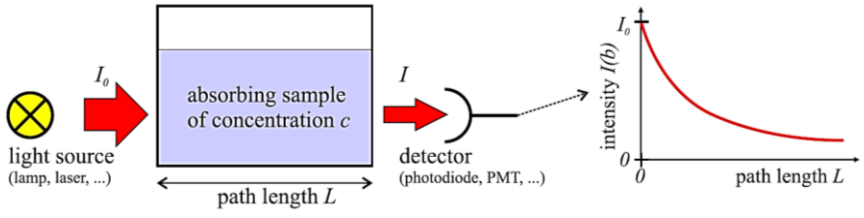


Figure 32. Set up of an absorption experiment.

$$A = \log_{10} \frac{I_0}{I}$$

Where I_0 and I are the intensities before after passing through the sample. In protein science, a very often used term for absorption is Optical Density (OD), for instance, if the OD=1 this means that only 10% of light passes through the cuvette (hence 90% is absorbed). The Beer-Lambert law is generally written as:

$$A = \varepsilon(\lambda) \cdot L \cdot [C]$$

Where $[C]$ is the sample concentration, L is the optical path length. $\varepsilon(\lambda)$ is a molar absorptivity and is a wavelength dependent coefficient expressed in $M^{-1}cm^{-1}$. For example, the molar absorptivity of the probe utilized in this study is:

Pyrromethene597 (ethanol): $\varepsilon(525) = 68000\ M^{-1}cm^{-1}$

Using Beer-Lambert law concentration can be easily assessed by the measurement of I_0 and I, given the path length and absorptivity of the molecule. Other way is plot absorption against a series of concentration and reading out the slope that gives $\varepsilon(\lambda) \cdot L$.

Absorption experiment also gives the possibility of identifying various components of the sample with absorptivity, for example DNA, proteins or fluorescent dyes. The measurement with absorption spectrometer (Figure 33) is absolute, meaning independent of the spectrometer and standing for a comparison between two measured intensities (thus reference sample is taken first to get rid of an influence of the solvent). [49]

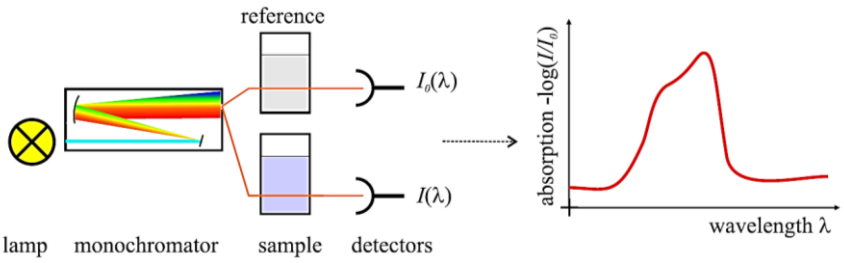


Figure 33. Scheme of an absorption experiment.

In this thesis, all absorption experiments were performed with Cary 300 Bio UV-Visible Spectrophotometer in quartz cuvettes (with cell path length 10 mm) in the wavelength range 200 – 800 nm (scan rate 600 nm/minute).

Calibration curve for pyrromethene597 was prepared in a concentration range 0.136 mM and 0.0136 mM (for the peak at 523 nm) and 0.1 mM – 0.005 mM for BSA (for the peak at 278 nm).

6.2.2 Fluorescence theory and experimental conditions

Fluorophores or fluorescent dyes (usually polyaromatic hydrocarbons or heterocycles) are exhibiting fluorescence as a result of a three-stage process in their electrons shell. This can be visualized with a simplified Jablonski diagram as in Figure 34.

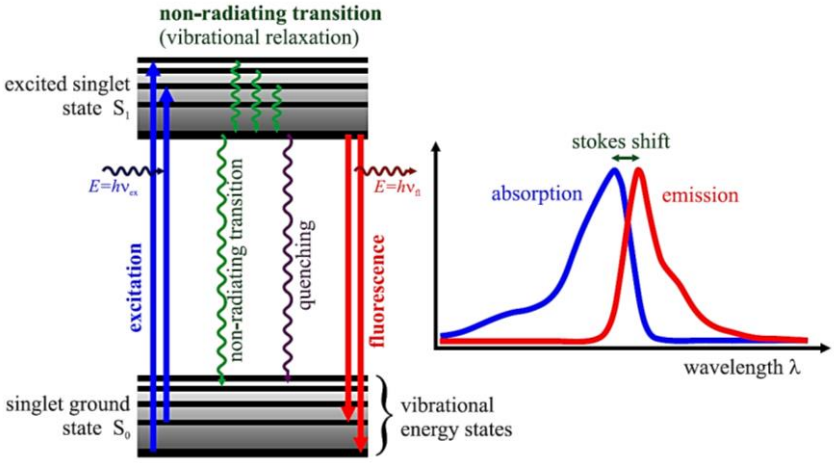


Figure 34. Simplified Jablonski diagram for the fluorescence emission with an exemplary spectra generated by combining absorption and emission spectra.

The first process is an excitation of the sample. The external light source (such as a laser) supplies a photon of an energy hV_{ex} and gets absorbed by a fluorophore. This energy pushes an electron from its ground state (S_0) to the excited state (S_1). Absorption takes place in around 1 fs (10^{-15} s). The second process is a non-radiating transition. On average the electron spends 1 – 10 ns in the excited state. In this time domain, the fluorophore can undergo conformational changes and can interact with its molecular environment. Non-radiative transition entails two major consequences. First, the energy of S_1 sub-states is dissipated to some extent giving rise to singlet excited state from which fluorescence comes. Second, not all molecules excited by a photon will reach back the ground state by fluorescence. This means there may be other processes involved such as collisional quenching, fluorescence energy transfer, and intersystem crossing and hence depopulate S_1 without a photon emitted. The last process is a fluorescence emission. The fluorophore can return to its ground state by emitting a photon of an energy hV_{fl} and reaching S_0 state. The energy of

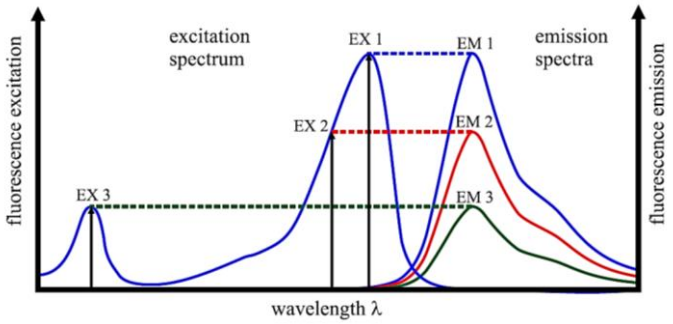


Figure 35. Excitation and emission spectra of a fluorophore. Spectra plotted at three wavelengths (EX1, EX2 and EX3).

such photon is lower due to energy dissipation during the excited state lifetime and thus has a longer wavelength than the excitation photon $\mathrm{hV_{ex}}$. This energy difference is directly related to so-called Stokes shift which can be expressed as a difference between absorption and emission maxima:

$$\Delta\lambda_{\mathrm{Stokes}} = \lambda_{\mathrm{max.emission}} - \lambda_{\mathrm{max.absorption}}$$

The fluorescence process is cyclic (unless the fluorophore is irreversibly damaged - photobleaching) meaning the fluorophore can be continually excited and detected. For the case of polyaromatic molecules in a solution, the distinct electronic transitions represented by $\mathrm{hV_{ex}}$ and $\mathrm{hV_{fl}}$ are swapped by broad energy spectra named fluorescence excitation spectrum and fluorescence emission spectrum (schematically represented in Figure 35). The width of each spectrum is of importance in applications that concern two or more fluorophores detected simultaneously.

With some uncertainty, the fluorescence spectrum of a single dye species in a solution is identical to its absorption spectrum (mirror-like). If applied conditions are the same the shape of fluorescence emission is independent of the excitation wavelength. The emission intensity is proportional to the amplitude of a fluorescence excitation spectrum at the excitation wavelength.

Fluorescence spectrometer consists of a monochromatized excitation light source and a detection channel (also with monochromator). The detection channel is placed at the angle of 90° to avoid the detection of transmitted light. Fluorescence spectrometer can be used in two modes:

- Detect emission spectra – where the excitation wavelength is fixed and detection takes place over a range of energies, resulting in the emission spectrum of the sample.
- Detect excitation spectra – where detection wavelength is fixed and the excitation energies are scanned – resulting in the excitation spectrum of the sample.[49]

The samples for the fluorescence measurement were performed on a Edinburgh F900 fluorometer. Samples prepared according to method 2 (protocol described in *6.1 Chemicals and sample preparation*).

6.2.3 Dynamic Light Scattering measurement

The experimental conditions of DLS measurement for BSA:BODYPY complex are the same than those used for GR and they are described in PROJECT1 methods.

6.2.4 Crystallization and Single Crystal X-ray Diffraction

Method description for crystallization and XRD analysis is reported in PROJECT1 methods. For crystallization of the protein:dye complex the conditions described in Sekula et al., [50] for pure BSA, were applied. There are two reservoir solutions are:

A: 0.2M calcium acetate, 0.1M MES (pH 6.5), 24% w/v PEG 5K;

B: 0.175M ammonium chloride, 0.1M MES (pH 6.5), 17% w/v MES 5K;

In addition 3 replicates were set with pure protein for eventual soaking experiments. The plate scheme is shown in Figure 36. Data were collected at ESRF Grenoble at ID23.1. The protein solution was maintained at a concentration of 66 mg/mL (according to the protocol). 2 µL of protein was placed on a siliconized coverslip and mixed with another 2 µL of mother solution from the reservoir (700 µL). Each well was vaseline-sealed and left in the incubator at 20 ± 1 °C. Control check was done every week under an optical microscope (Leitz Wetzlab SM-LUX) equipped with a digital camera Motic 5.0 MP.

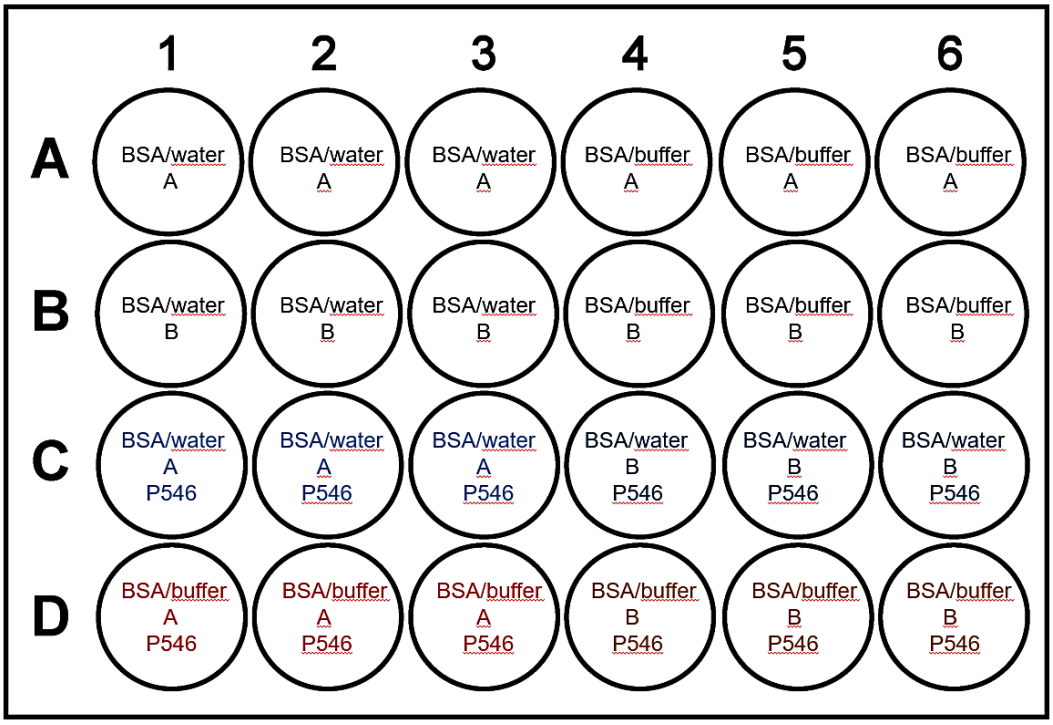


Figure 36. Plate for hanging drop vapor diffusion crystallization of BSA and pyrromethene complex. A and B indicates two different crystallization conditions taken from literature [50]. P597 indicates pyrromethene597 as a supposed adduct with the protein. Composition of reservoir solutions: A – 0.2M calcium acetate, 0.1M MES (pH 6.5), 24% w/v PEG 5K Da, B – 0.175M ammonium chloride, 0.1M MES (pH 6.5), 17% PEG 5K Da.

Chapter 7 - Results and discussion

Fluorescent probes have always been direct ‘tools’ in achieving visual results in live cell imaging. The physicochemical characteristics of the probe determine the quality of the image and very often it is a matter with certain difficulties. This work is based on pyrromethene family of fluorescent dyes (BODIPY) as a robust probe in molecular biology, due to its excellent properties (see introduction). A cell as a fragile target needs a lot of care while handling, thus any additional procedure (such as staining) has to be reasonably designed. From BODIPY dyes group only a handful of water soluble molecules are available as reported by Loudet *et al.* (2007).

One of the possible solutions is an incorporation of the hydrophobic probe into an easily available protein, such as lysozyme or BSA. The advantage of developing steady protocols is that the model proteins are water soluble, thermally stable and relatively resistant to the harsh chemical environment. This project is an attempt in achieving the stable protein-dye adduct for imaging purposes. Once the preparatory step was settled optical spectroscopies were employed to assess the quality and efficiency of the protocol.

7.1 Absorption experiments

BSA is difficult to crystallize protein and number of operations done to incorporate the probe makes it even more complex target for crystallographic study. Thus techniques such as absorption and fluorescence emission can be employed to get useful conclusions about the system. As a first try absorption experiments were performed to get an idea how the protein adduct behave. The main issue to overcome is a strongly hydrophobic character of the probe thus excessive concentration was utilized to see whether it has a positive influence on the absorption profile of the protein band. Before the tryouts, pyrromethene597 methanol solution was taken for calibration curve preparation for further quantification. This dye has the highest solubility in methanol but adduct preparation is water-based due to the nature of the protein. Solubility limit for the dye is around 0.5 mM (in methanol) and the calibration curve with respective absorption spectra is shown in Figure 37. The calibration curve was prepared on absorption maxima of pyrromethene597 thus 523 nm.

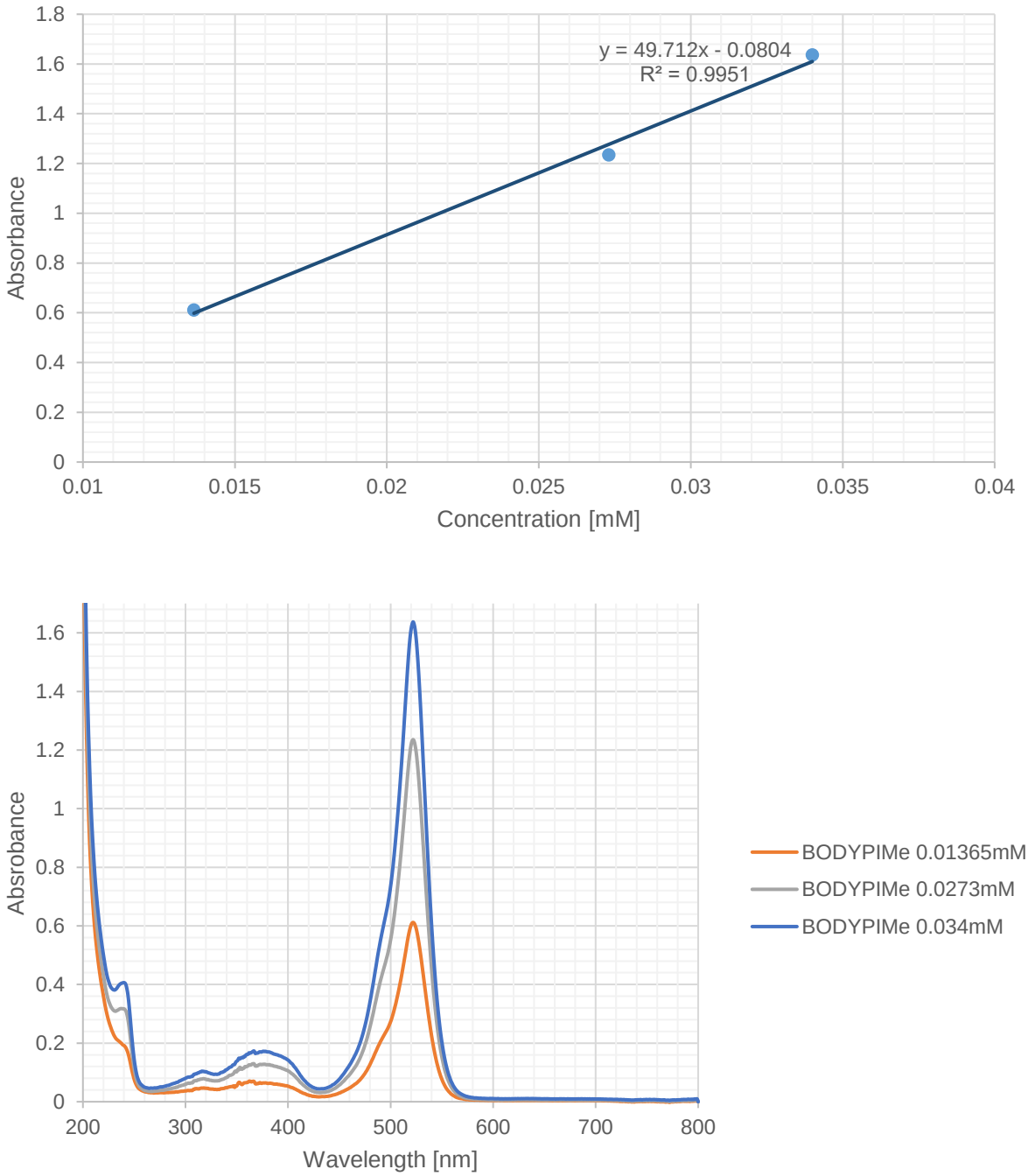


Figure 37. Top: Calibration curve for pyrromethene597 absorption maxima at 523 nm. Bottom: Absorption profiles of pyrromethene597 at different concentrations (dye dissolved in methanol).

From the slope of the calibration curve molar extinction coefficient (ε) can be calculated. The theoretical value of ε (in ethanol) for pyrromethene597 is equal to 68000 M^{-1} cm^{-1} nevertheless in the case of a spectrometer used in this work absorption maxima is observed at 523 nm and ε (in methanol) is equal to 49712 M^{-1} cm^{-1}. This difference presumably comes from the instrument calibration, solvent type, and weight accuracy.

Figure 38A presents absorption profile in the range 200-800 nm for the protein reference and BSA:pyrromethene597. The low intensity of the absorption band of BODIPY indicates that magnetic stirring leads to formation of aggregates of BODIPY and the broad shape of the band indicates that the dye is not present in a monomeric form. In fact, pyrromethene597 is highly absorbing at 523 nm and the band appear very sharp. Figure 38B presents size distribution in both samples with DLS. Size distribution of intensity is not very conclusive in this case, likely due to the aggregates in both cases dominating in the solution.

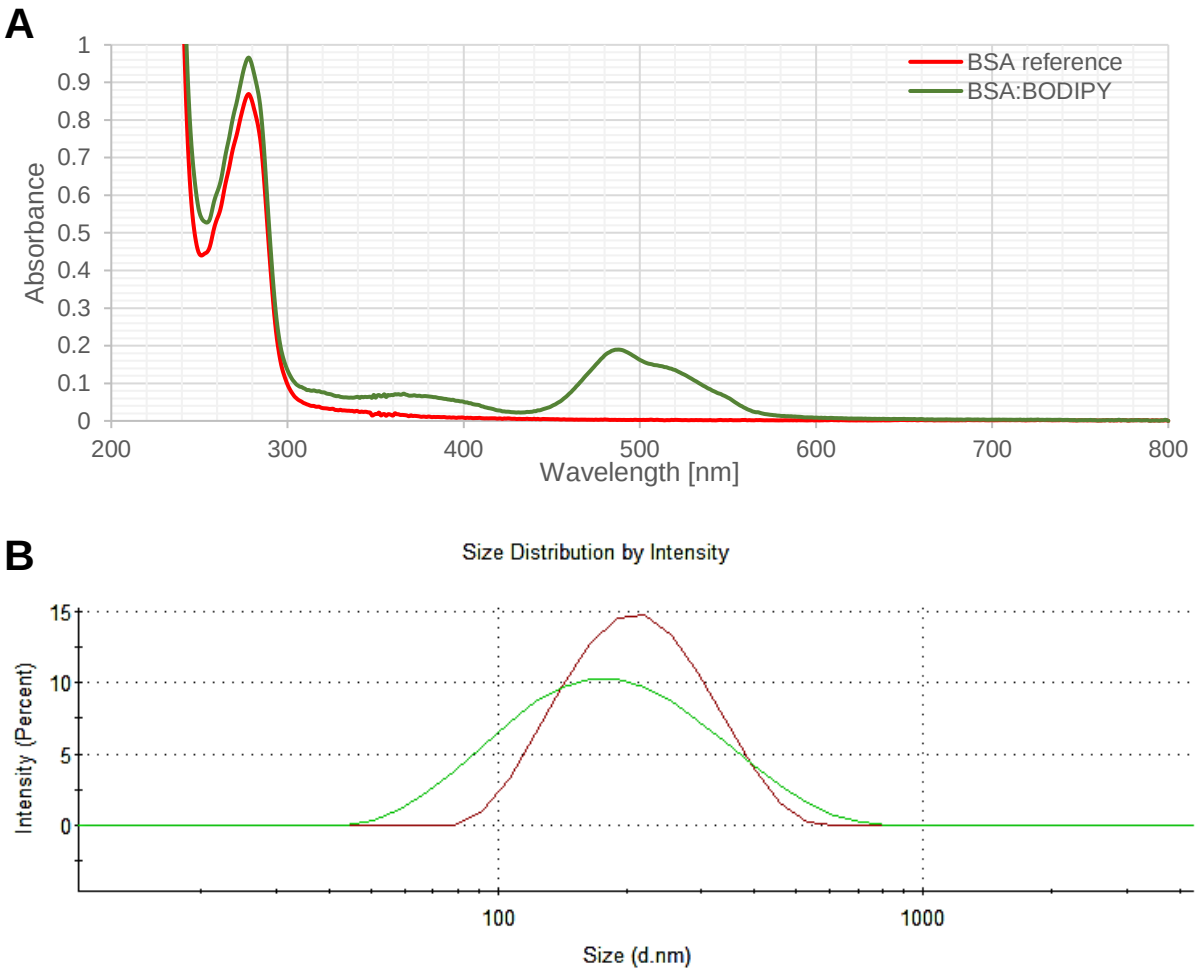


Figure 38. (A) UV-Vis spectra of the protein reference and an adduct of BSA:pyrromethene597. Sample diluted 1:10 for the experiment (thus final concentration of the protein was 0.025mM and pyrromethene597 was 0.55mM). (B) DLS size distribution by intensity, where red spectra is a BSA reference while green corresponds to BSA:pyrromethene597 adduct.

After centrifugation, there was a significant amount of pellet (mainly of pyrromethene597) that was furtherly quantified by recovery with methanol. The starting amount of the dye was 1.5 mg and from the pellet (using calibration curve) 1.39 mg was recovered, meaning

that very little amount was in fact in the solution. Thus, starting ratio between BSA and pyrromethene597 was 0.5 mM : 5.72 mM and the final ratio (thus after stirring and centrifugation) was 0.493 mM : 1.38 mM.

Next concept was to investigate different dye molecule incorporation methodology. As the aggregation is pronounced, microtip sonication was applied to see if this could anyhow improve the condition. For this idea two mixing methods were utilized to check the effectiveness, one sample (protein:dye ratio 1:12) left under magnetic stirring and the second (same ratio) taken for sonication. UV-Vis spectra were taken after every 60 minutes of magnetic stirring (MS) and every 10 minutes of microtip sonication (S) for 3 to 5 cycles respectively. After the experiment, sonicated sample was left at 4 °C for 24 hours to check the stability of the solution (Figure 39A). MS sample was left under stirring for 24 hours to see if better results are obtained when protein:dye ratio is lowered compared to the previous case of 1:20 (Figure 39B).

Indeed sonication was found to be much more efficient at dissolving pyrromethene597 to a visible extent. Less than an hour was enough to see the evident difference in an absorption profile. More importantly, a radical change in pyrromethene597 absorption band was recorded. The main peak was blue shifted by around 30 nm and did not maintain its spiky shape (as in calibration curve spectra).

Decreasing the ratio from 1:20 to 1:12 worked better for MS too. Absorption maxima for pyrromethene597 increased from 0.2 up to 1.8 (just for different ratios) after 24 hours of MS. Still, maxima are less intense than in the sonicated sample but peak shifting and shape change is maintained. After centrifugation, DLS was taken again, but as in the previous method, only large aggregates are visible (~150 nm for sonicated sample and ~250 nm for the sample under magnetic stirring). Similarly, pellet after centrifugation was quantified and from the starting amount of 1.6 mg of the dye, 1.1 mg was recovered meaning this incorporation method was more efficient than the first one. The starting ratio of BSA and pyrromethene597 was 0.25 mM : 3.11 mM, while after sonication and centrifugation ration decreased to 0.248 mM : 1.88 mM.

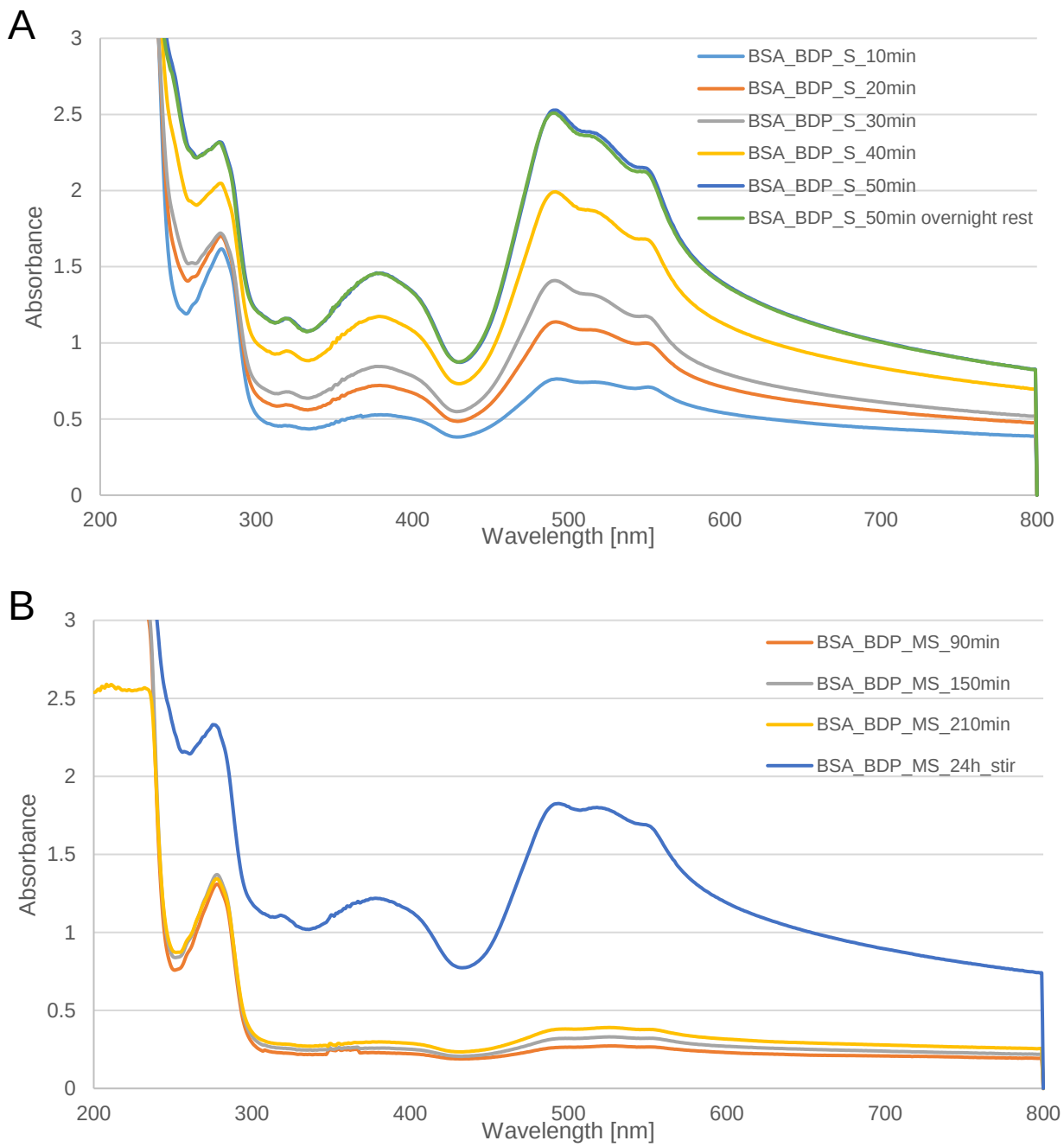


Figure 39. (A)UV-Vis profile of protein:dye adduct prepared with microtip sonication. Sample was sonicated for 50 minutes in total and spectra was acquired every 10 minutes. The green spectra stands for the sample analyzed after 24 hours 'rest' at 4 °C. (B) UV-Vis profile of protein:dye adduct left under magnetic stirring. Spectra acquired every hour and left for 24 hours at room temperature.

From the absorption and DLS experiments it is possible to hypothesize that BODIPY form, in the aqueous protein solution, aggregates with a hydrodynamic radius bigger than 50 nm in the case of sonication and 100 nm in the case of magnetic stirring, surrounded and stabilized by several BSA molecules. Since BODIPY alone in water solution form bigger and irregularly sized aggregates (Figure 40; Hydrodynamic radius about 522 nm) it can be concluded that the BSA has a major role in stabilizing BODIPY particle in solution preventing high aggregation.

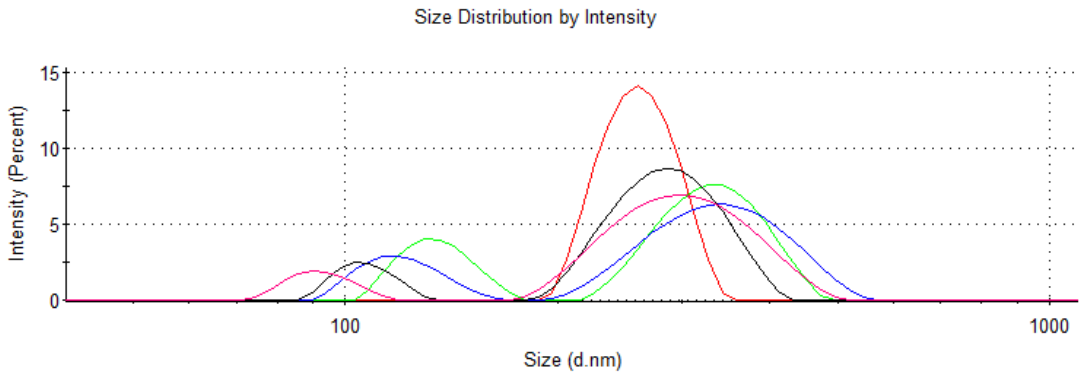


Figure 40. DLS profile of BODIPY water solution (0.04 mM) prepared with microtip sonication (60 minutes). Size distribution by intensity indicates irregular aggregates in the range between 80 - 522 nm.

7.2 Fluorescence experiments

After absorption experiments, we have performed emission spectroscopy measurements to gain an insight on molecular and aggregated BODIPY in the BSA solution. Emission spectra revealed that the emission of BODIPY does not show significant variations compared with BODIPY in methanol. For this fluorescence lifetime, experiments provided some basic information about their photophysical specificity. In addition to the previous protocols with magnetic stirring and sonication, a separation step was employed – viva-spin. The sample as being at unknown state, the viva-spin separation was utilized with the pore size suitable for separation of proteins up to 100 kDa. This in fact eliminated the larger aggregates. Since sonication was found to be more efficient, MS samples were not taken for further consideration. Figure 41 presents fluorescence emission from two samples: sonicated (protein:dye at ratio ~1:10) and the same sample taken for viva-spin separation (10 minutes, 10000 rpm at 4 °C). Excitation spectra confirmed indeed that the

observed emission only comes from monomeric BODIPY (thus featuring the same absorption spectrum as BODIPY in methanol). On the contrary the additional band observed in the absorption spectrum does not lead lead to emission. This allows for identification of those bands as the bands of not emissive aggregates dispersed in a solution, while the little emission might arise from monomeric BODIPY embedded inside BSA protein.

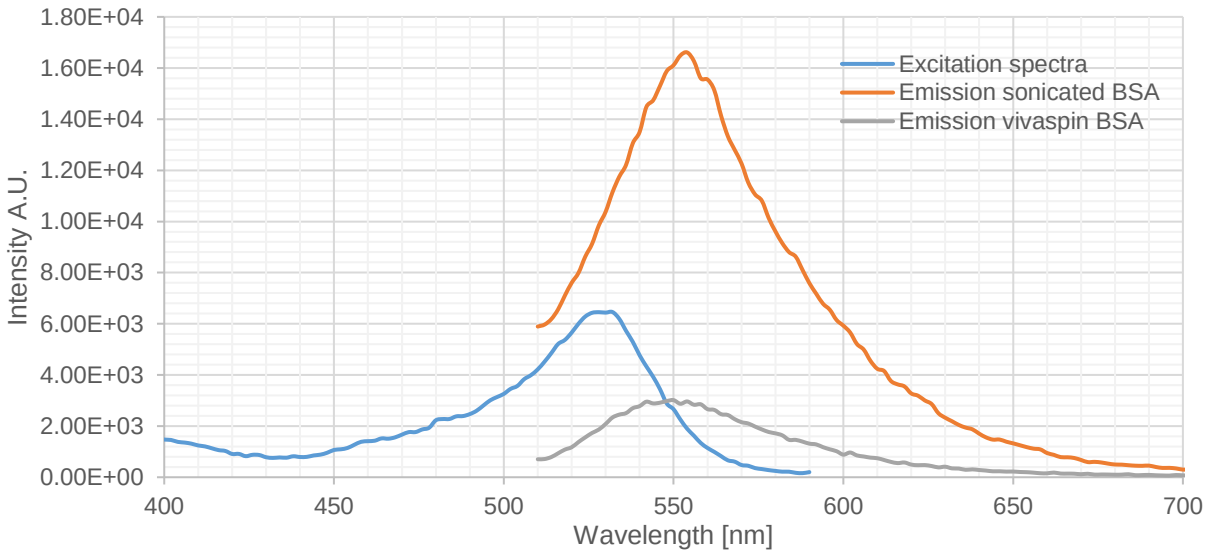


Figure 41. Fluorescence spectra for protein:dye complex. The orange spectrum is a sample taken for 50 minutes sonication, while the grey one was additionally taken for viva-spin separation.

Separation procedure significantly reduced the number of aggregates as observed in fluorescence lifetime for both samples (Figure 42). From fluorescence emission, it is seen that concentration is drastically reduced for the sample with a separation step.

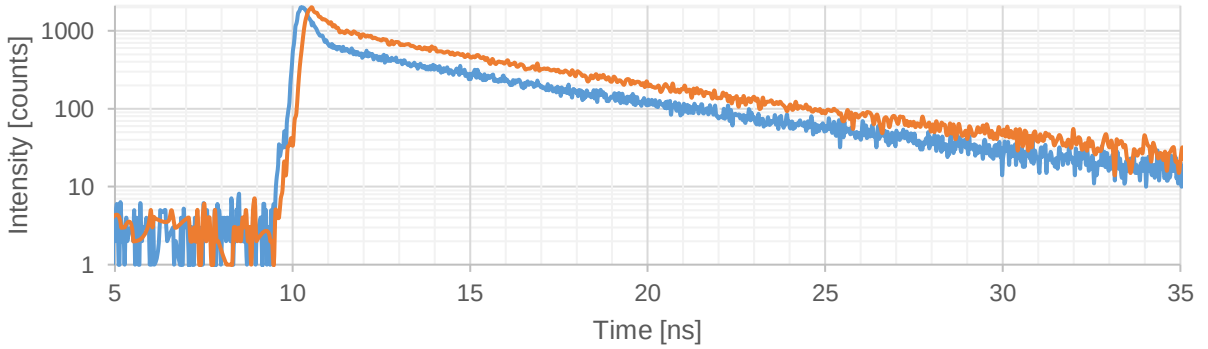


Figure 42. Fluorescence lifetime for sonicated adduct (blue) and sonicated/separated sample (orange).

Just from the fluorescence lifetime profile, it is seen that sonicated/separated sample has a lower contribution of one exponential component. For both spectra the fit has two exponential components, indicating two different emissions 'characters'. The fitting was done in software FAST for fluorescence emission data processing, according to the formula for the double-exponential decay of fluorescence intensity:

$$I(t) = A_1 \exp\left(-\frac{t}{\tau_1}\right) + A_2 \exp\left(-\frac{t}{\tau_2}\right)$$

Where

$I(t)$ – fluorescence intensity,

A_1, A_2 – pre-exponential factors,

τ_1, τ_2 – lifetimes,

t – pulse time.

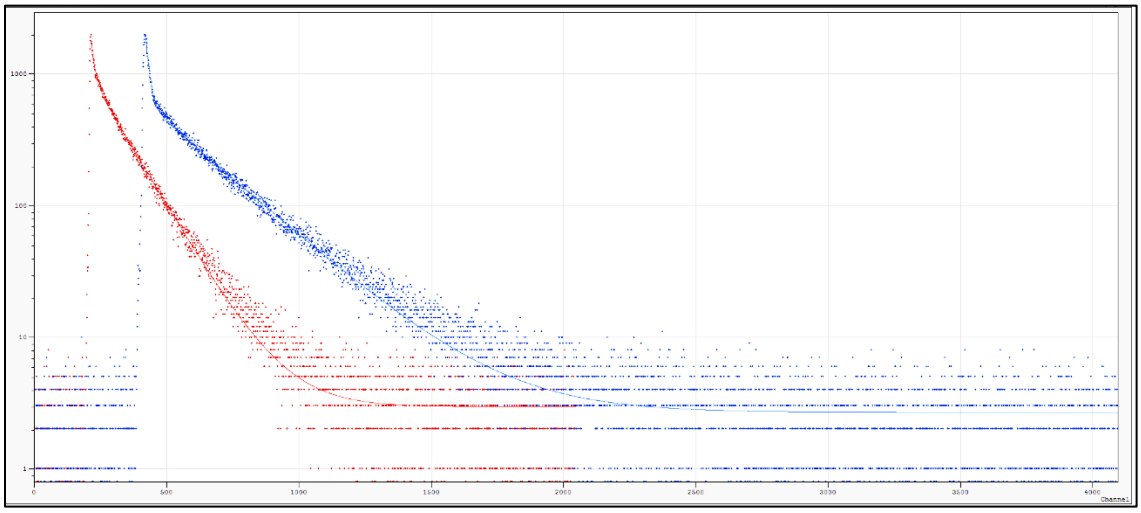


Figure 43. Fluorescence lifetime for sonicated protein:dye solution (blue) and sonicated and viva-spin separated solution (red).

Fittings with superimposed spectra for both samples is shown in Figure 43. Table 6 presents fitting parameters calculated in the software FAST.

Table 6. Fitting parameters for sonicated and sonicated/separated sample.

Sample type	A_1	A_2	τ_1 [ns]	τ_2 [ns]
Sonicated sample	1612.54	605.5	0.341	6.079
Sonicated/separated sample	949.8	947.14	0.644	6.23

Last, but not least fluorescence measurement was the lifetime analysis at different emission wavelengths. Figure 44 presents measurements where the excitation wavelength was 405 nm and emission lifetime measured in the range 520 – 700 nm. This measurement was done for sonicated/separated sample. Even though the number of aggregates was supposed to be decreased, the overlapped spectra show a clear contribution of a differently emitting form of pyrromethene597. For the wavelengths of 520 and 540 nm, the profile is maintained as in Figure 43 (red line) thus 'more emitting' structure. From 560 nm and on, the emission undoubtedly changes its behavior, thus probably it is the energy range where the separate form of pyrromethene597 is emitting (supposed aggregates and/or nanoparticles).

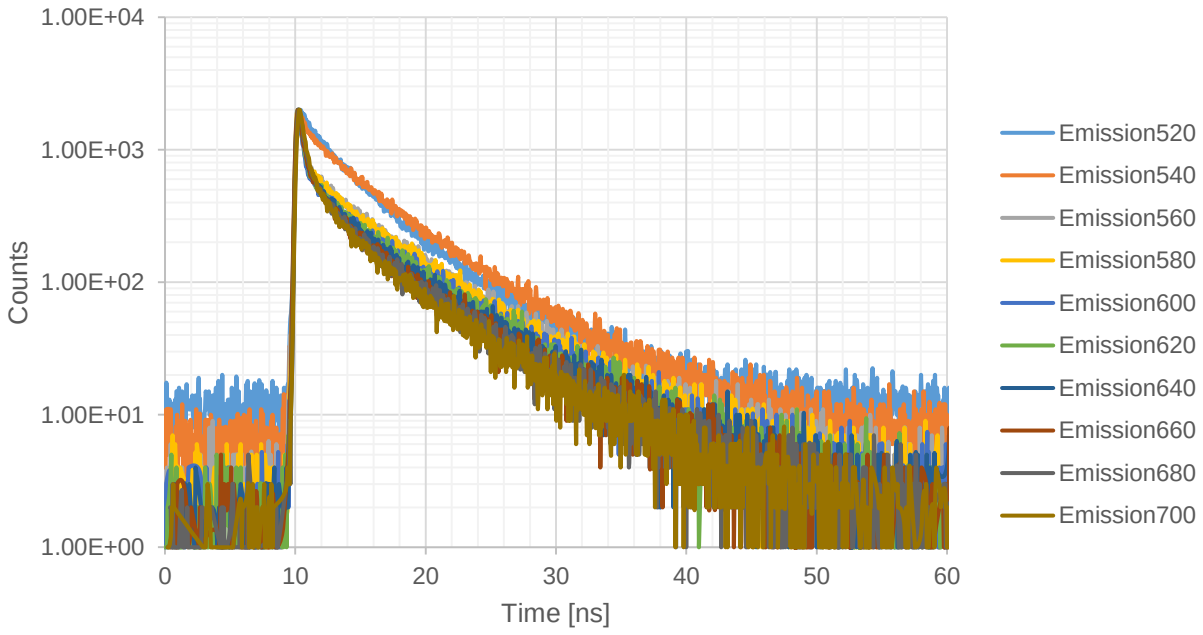


Figure 44. Emission profiles of sonicated/separated protein:dye sample. Excitation wavelength was equal to 405 nm and emission lifetime measured in the range from 520 nm up to 700 nm.

7.3 Crystallographic study

BSA crystal structure was resolved quite recently (2012) and the best resolution found on PDB is 2.47 Å. Thus without trials, the methodology was reproduced for this study. Bujacz *et al*.(2012) reported two reservoir solutions suitable for BSA crystallization

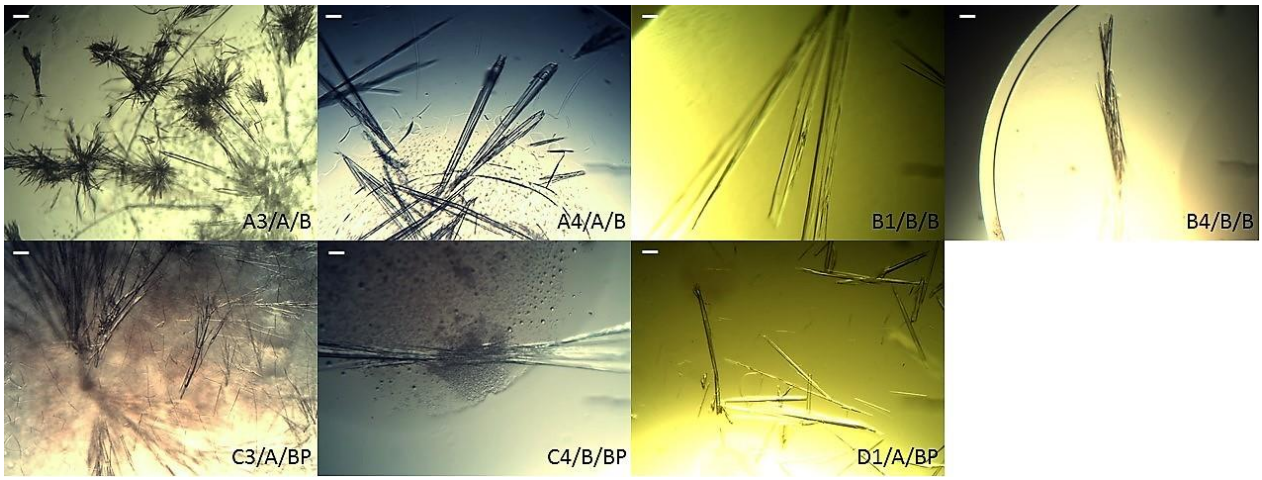


Figure 45. Images of crystallization positive hints. Naming scheme: well#/crystallization condition/sample (where B stands for BSA and BP for BSA:pyrromethene597). Scale bar in the top left corner indicates 70 µm.

(composition in *6.2.4 Crystallization and Single Crystal X-ray Diffraction*). According to this protocol crystals should be formed after seven days. In this case, it was a condition difficult to reproduce and after one week, crystals were observed only for BSA:pyrromethene597 adduct (luckily). In addition (as shown in methodology chapter) BSA was crystallized in a pure form (in water or buffer) for eventual soaking trials. After 3 months few crystals of pure BSA were observed, meaning the conditions are right but the time of crystal growth is not reproducible. Figure 45 shows results of the BSA and BSA:pyrromethene597 crystallization. After one week crystals for BSA:pyrromethene597 were observed (thus conditions C3/A/BP, C4/B/BP, and D1/A/BP). Those were taken for XRD analysis at ESRF Grenoble (ID 23.1 beamline). One peculiarity spotted was no influence of buffer in the protein sample, at least no visible trend. On Figure 45 condition A3 and B1 show BSA in water while A4 and B4 is a BSA sample in the buffer. Thus morphology is affected by the reservoir solution and not buffer addition in the protein preparation step.

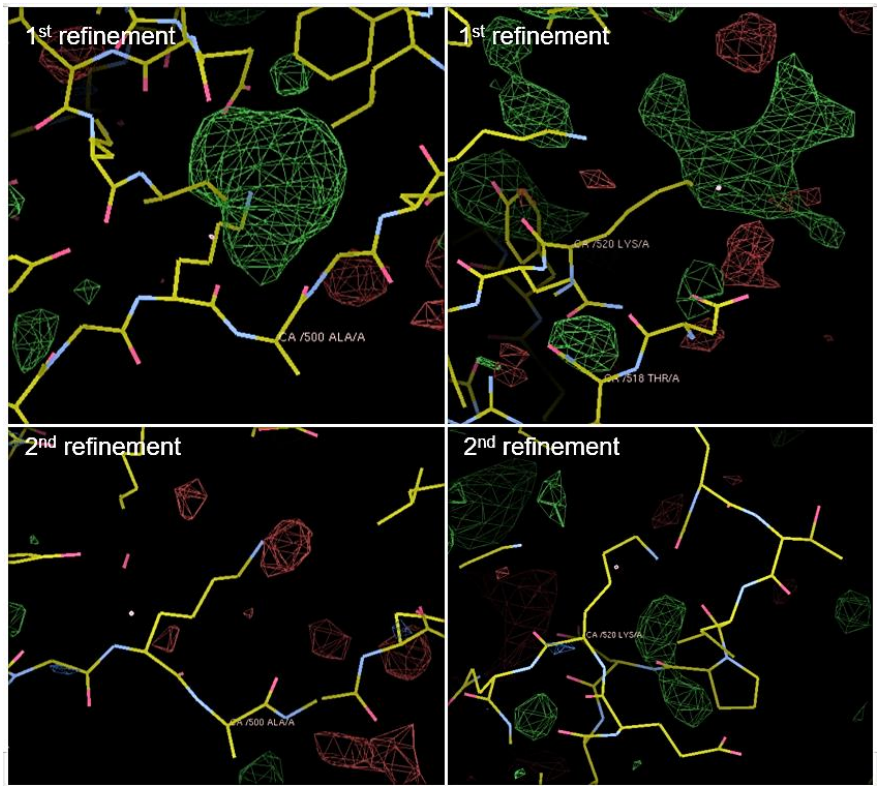


Figure 46. Example of pyrromethene597 fitting and subsequent refinement cycle. Only experimental density is shown (F_0).

Data collection, phase problem solution, and model building procedure were just like in PROJECT1. The structure solution was definitely less time-consuming due to BSA model availability and sequence identity of 100%. Data collection details are presented in Table 7. During model building, certain densities were identified that could be assigned as a pyrromethene597 molecule (Figure 46 top). Then a BODIPY molecule was built inside, but unfortunately in the subsequent refinement cycle, the density disappeared. This phenomenon can be due to a low occupancy of the dye in the protein binding sites or to a local disorder of the dye, both cases producing a weak electron density map preventing correct model building and structure refinement.

Table 7. Data collection parameters for BSA:pyrromethene597 crystals.

Parameter	BSA:pyrromethene597
Data collection	
Unit cell	a=216.8 Å, b=44.4 Å, c=140.3 Å, α= 90°, β=114.1°, γ=90°
Space group	C 1 2 1
Resolution range* (Å)	98.942 – 3.039 Å (3.091 – 3.039 Å)
Unique reflections	1691
Completeness* (%)	99.8 (99.5)
R_{merge}*	5.2 (78.0)
I/σ	8.9 (2.1)
Multiplicity	6.6 (6.8)

7.4 Conclusions

Optical spectroscopies were found to be useful tools in understanding behavior of the protein:dye system. Hydrophobic nature of the fluorescent probe turned out to be more complex in a protein solution than expected. Absorption and fluorescence experiments were successfully employed in the primary assessment of what might be happening during sample preparation. Formation of aggregates with irregular dynamic performance is somewhat unavoidable. Stability of the system can be progressed by applying sonication step in the adduct preparation protocol.

One powerful method to investigate whether pyrromethene597 is incorporated into BSA structure would be mass spectrometry analysis. This could possibly give a rapid idea if in the solution protein adduct is present, thus protein molecular weight increased by the mass of pyrromethene597 as a molecular ion peak. This approach was taken into consideration from the early stages of the project, however, due to instrument availability limitation, it was not possible until the thesis submission deadline.

Having grown crystals of BSA also trials with soaking in pyrromethene597 solution will be considered in nearby future. Another approach will include the application of various (and commercially available) pyrromethene probes and eventually utilization of other model proteins (such as lysozyme).

References

[1] Deponte, M. (May 2013). Glutathione catalysis and the reaction mechanisms of glutathione-dependent enzymes. Biochim. Biophys. Acta., 1830 (5).

[2] Mukwevho, M., Ferreira, Z., Ayeleso, A. (2014). Potential Role of Sulfur-Containing Antioxidant Systems in Highly Oxidative Environments. Molecules, 19(12), 19376-19389.

[3] Deponte, M., Hell, K. (2009). Disulphide bond formation in the intermembrane space of mitochondria. J. Biochem., 146, 599–608.

[4] Schulz, G.E., Schirmer, R.H., Sachsenheimer, W., Pai, E.F. (1978). The structure of the flavoenzyme glutathione reductase. Nature, 273 (5658).

[5] Thieme, R., Pai, E.F., Schirmer, R.H., Schulz, G.E. (1981). Three-dimensional structure of glutathione reductase at 2 Å resolution. J. Mol. Biol. 152, 763–782.

[6] Deponte, M., Urig, S., Arscott, L.D., Fritz-Wolf, K., Reau, R., Herold-Mende, C., Koncarevic, S., Meyer, M., Davioud-Charvet, E., Ballou D.P., Williams Jr., C.H., Becker, K. (2005) Mechanistic studies on a novel, highly potent gold-phosphole inhibitor of human glutathione reductase. J. Biol. Chem., 280, 20628–20637.

[7] Outten, C.E., Culotta, V.C. (2004). Alternative start sites in the *Saccharomyces cerevisiae* GLR1 gene are responsible for mitochondrial and cytosolic isoforms of glutathione reductase. J. Biol. Chem., 279, 7785–7791.

[8] Imlay, J.A. (1983). Pathways of oxidative damage. Annu. Rev.Microbiol., 57, 395–418.

[9] Pai, E.F., Schulz, G.E. (1983). The catalytic mechanism of glutathione reductase as derived from X-ray diffraction analyses of reaction intermediates. J. Biol. Chem., 258, 1752–1757.

[10] Huber, P.W., Brandt, K.G. (1980). Kinetic studies of the mechanism of pyridine nucleotide dependent reduction of yeast glutathione reductase. Biochemistry, 19, 4569–4575.

[11] Krauth-Siegel, R.L., Arscott, R.D., Schonleben-Janas, A., Schirmer, R.H., Williams Jr., C.H. (1998). Role of active site tyrosine residues in catalysis by human glutathione reductase. Biochemistry, 37, 13968–13977.

[12] Wong, K.K., Vanoni, M.A., Blanchard, J.S. (1988). Glutathione reductase: solvent equilibrium and kinetic isotope effects. Biochemistry, 27, 7091–7096.

[13] Merchant, Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A. et al. (2007). The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions. Science, 318 (5848).

[14] Malvern Instruments: Dynamic Light Scattering – An introduction in 30 minutes, technical note.

[15] Hampton research website, https://hamptonresearch.com/documents/growth_101/3.pdf, access on 15.05.2017

[16] McPherson, A., DeLucas, L.J. (2015). Microgravity protein crystallization, npj Microgravity 1, 15010

[17] Garman, E. F., Schneider, T. R. (1997). Macromolecular Cryocrystallography. Journal of Applied Crystallography, 30 (3).

[18] Hauptman, H. (1997). Phasing methods for protein crystallography. Curr. Opin. Struct. Biol. 7 (5).

[19] Jeruzalmi, D. (2006). First analysis of macromolecular crystals: biochemistry and x-ray diffraction. Methods Mol. Biol., 364: 43–62.

[20] Jerrold, B., Seibert, A., Leidholdt, E. (2012). The Essential Physics of Medical Imaging. USA: Lippincott Williams & Wilkins, p. 116.

[21] Veksler, V. I. (1983). A new method of accelerating relativistic particles. Comptes Rendus de l'Academie Sciences de l'URSS 43 (8): 329–331.

[22] Romaniuk, R.S., Koutchouk, J.P. (2010). Determination of beam intensity and position in a particle accelerator; Editorial Series on Accelerator Science and Technology, Vol.11.

[23] ESRF website, http://www.esrf.eu/about/synchrotron-science/synchrotron, accessed on 12.04.2017.

[24] Winn et al. (2011). Overview of the CCP4 suite and current developments. Acta. Cryst., D67, 235-242.

[25] Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. (2010). Features and development of Coot. Acta Cryst. D66, 486-501.

[26] Evans, P.R. (2006). Scaling and assessment of data quality. Acta Cryst. D62, 72-82.

[27] Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. J. Appl. Cryst. 21, 916-924.

[28] French, G.S., Wilson K.S. (1978). On the treatment of negative intensity observations. Acta. Cryst. A34, 517.

[29] Monaco, S., Gordon, E., Bowler, M., Delagenière, S., Guijarro, M., Spruce, M., Spruce, D., Svensson, O., McSweeney, S.M., McCarthy, A.A., Leonarda, G., Nanaob, M.H. (2013). Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF. J. Appl. Cryst. 46, 804-810.

[30] Yu, J., Zhou, C.Z. (2007). Crystal structure of glutathione reductase Glr1 from the yeast *Saccharomyces cerevisiae*. Proteins, 68(4):972-9.

[31] Murshudov, G., Vagin, A., Dodson, E. (1996). Application of Maximum Likelihood Refinement. Refinement of Protein structures, Proceedings of Daresbury Study Weekend.

[32] De Colibus, L., Mattevi, A. (2006). New frontiers in structural flavoenzymology. Current Opinion in Structural Biology,16(6):722–728.

[33] Berkholz, D. S., Faber, H. R., Savvides, S. N., Karplus, A. (2008). Catalytic cycle of human glutathione reductase near 1 Å resolution. Journal of Molecular Biology, 382(2):371-84.

[34] Pace, R. N., (2006). Time for change. Nature, 441: 289.

[35] Ettinger, A., Wittmann, T. (2015). Fluorescence Live Cell Imaging. Methods Cell Biology, 123: 77–94.

[36] Johnson, I. (1998). Review: Fluorescent probes for living cells. The Histochemical Journal, 30(3): 123–140.

[37] Valeur, B., Berberan-Santos, M. N. (2012). Molecular Fluorescence: Principles and Applications. Wiley-VCH, 2nd edition.

[38] Ziessel, R. (2007). The chemistry of Bodipy: A new *El Dorado* for fluorescence tools. New Journal of Chemistry (4).

[39] Treibs, A., Kreuzer. F. H. (1968). Difluorboryl-Komplexe von Di- und Tripyrrylmethenen. Justus Liebigs Annalen der Chemie, 718 (1): 208-223.

[40] Schmitt, A., Hinkeldey, B., Wild, M., Jung, G. (2009). Synthesis of the Core Compound of the BODIPY Dye Class: 4,4′-Difluoro-4-bora-(3a,4a)-diaza-s-indacene. J. Fluoresc., 19 (4): 755–759.

[41] Tram, K., Yan, H., Jenkins, H. A., Vassiliev, S., Bruce, D. (2009). The synthesis and crystal structure of unsubstituted 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene (BODIPY). Dyes and Pigments, 82 (3): 392–395.

[42] Arroyo, I. J., Hu, R., Merino, G., Tang, B. Z., Peña-Cabrera, E. (2009). The Smallest and One of the Brightest. Efficient Preparation and Optical Description of the Parent Borondipyrromethene System. J. Org. Chem., 74 (15): 5719–22.

[43] Thoresen, L. H., Kim, H., Welch, M. B., Burghart, A., Burgess, K. (1998). Synlett, 1276.

[44] Thermo fisher website, https://www.thermofisher.com/it/en/home/references/molecular-probes-the-handbook/fluorophores-and-their-amine-reactive-derivatives/bodipy-dye-series.html, accessed on 03.05.2017

[45] Yamada, K.; Toyota, T.; Takakura, K.; Ishimaru, M.; Sugawara, T. (2001). Preparation of BODIPY probes for multicolor fluorescence imaging studies of membrane dynamics. New Journal of Chemistry, 25, 667.

[46] Nicolaou, K. C., Claremon, D. A., Papahatjis, D. P. (1981). Tetrahedron Letters, 22, 4647.

[47] Loudet, A., Burgess, K. (2007). BODIPY Dyes and Their Derivatives: Syntheses and Spectroscopic Properties. Chem. Rev., 107, 4891-4932.

[48] Sugio, S.; Kashima, A.; Mochizuki, S.; Noda, M.; Kobayashi, K. (1999). Crystal structure of human serum albumin at 2.5 A resolution. Protein Engineering Design and Selection, 12 (6): 439–446.

[49] Toth, K., Langowski, J. Biophysics of Macromolecules (B040), Absorption and Fluorescence Spectroscopy, German Cancer Research Center (DKFZ).

[50] Sekula, B., Zielinski, K., Bujacz, A. (2013). Crystallographic studies of the complexes of bovine and equine serum albumin with 3,5-diiodosalicylic acid. Int. J. Biol. Macromol. 60C, 316-324.

[51] Madhu, S., Gonnade, R., Ravikanth, M. (2013). Synthesis of 3,5-Bis(acrylaldehyde) Boron-dipyrromethene and Application in Detection of Cysteine and Homocysteine in Living Cells. The Journal of Organic Chemistry, 78, 5056−5060.

Additional information

1. Reagent solution lists for crystallization trials optimization:
    - Manually prepared reagent solutions (based on trials from MD1-01-CF)
    - Crystal Extension kit (Crystal Screen 2) – Hampton Research
    - JCSG-plus MD1-37 for soluble proteins - Molecular Dimensions

Manual reagent solutions

Tube#	Salt	Buffer	pH	Precipitant
18A	0.2M sodium acetate	0.1M MES	6.5	18% PEG 8K
18B	0.2M sodium acetate	0.1M MES	6.5	18% PEG 4K
18C	0.2M magnesium chloride	0.1M MES	6.5	18% PEG 8K
18D	0.2M magnesium chloride	0.1M MES	6.5	18% PEG 4K
18E	0.2M sodium acetate	0.1M HEPES	7.5	18% PEG 8K
20A	none	0.1M MES	6.5	18% PEG 8K
20B	0.2M calcium acetate	0.1M HEPES	7.5	18% PEG 8K
31A	none	0.1M HEPES	7.5	20% PEG 4K
33A	0.2M magnesium chloride	0.1M Tris	8.5	25% PEG 4K
33B	0.2M magnesium chloride	0.1M Tris	8.5	20% PEG 4K
33C	none	0.1M Tris	8.5	30% PEG 4K
33D	0.2M calcium acetate	0.1M Tris	8.5	30% PEG 4K
33E	0.2M ammonium acetate	0.1M Tris	8.5	30% PEG 4K
33F	none	0.1M Tris	8.5	18% PEG 8K
33G	0.2M calcium acetate	0.1M Tris	8.5	18% PEG 8K
33H	0.2M ammonium acetate	0.1M Tris	8.5	18% PEG 8K

Crystal Screen 2™ — HR2-112 Reagent Formulation

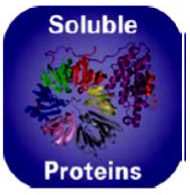
Tube #	Salt	Buffer ◊	Precipitant
1.	2.0 M Sodium chloride	None	10% w/v Polyethylene glycol 6,000
2.	0.5 M Sodium chloride 0.01 M Magnesium chloride hexahydrate	None	0.01 M Hexadecyltrimethylammonium bromide
3.	None	None	25% v/v Ethylene glycol
4.	None	None	35% v/v 1,4-Dioxane
5.	2.0 M Ammonium sulfate	None	5% v/v 2-Propanol
6.	None	None	1.0 M Imidazole pH 7.0
7.	None	None	10% w/v Polyethylene glycol 1,000 10% w/v Polyethylene glycol 8,000
8.	1.5 M Sodium chloride	None	10% v/v Ethanol
9.	None	0.1 M Sodium acetate trihydrate pH 4.6	2.0 M Sodium chloride
10.	0.2 M Sodium chloride	0.1 M Sodium acetate trihydrate pH 4.6	30% v/v (+/-)-2-Methyl-2,4-pentanediol
11.	0.01 M Cobalt(II) chloride hexahydrate	0.1 M Sodium acetate trihydrate pH 4.6	1.0 M 1,6-Hexanediol
12.	0.1 M Cadmium chloride hydrate	0.1 M Sodium acetate trihydrate pH 4.6	30% v/v Polyethylene glycol 400
13.	0.2 M Ammonium sulfate	0.1 M Sodium acetate trihydrate pH 4.6	30% w/v Polyethylene glycol monomethyl ether 2,000
14.	0.2 M Potassium sodium tartrate tetrahydrate	0.1 M Sodium citrate tribasic dihydrate pH 5.6	2.0 M Ammonium sulfate
15.	0.5 M Ammonium sulfate	0.1 M Sodium citrate tribasic dihydrate pH 5.6	1.0 M Lithium sulfate monohydrate
16.	0.5 M Sodium chloride	0.1 M Sodium citrate tribasic dihydrate pH 5.6	2% v/v Ethylene imine polymer
17.	None	0.1 M Sodium citrate tribasic dihydrate pH 5.6	35% v/v tert-Butanol
18.	0.01 M Iron(III) chloride hexahydrate	0.1 M Sodium citrate tribasic dihydrate pH 5.6	10% v/v Jeffamine® M-600®
19.	None	0.1 M Sodium citrate tribasic dihydrate pH 5.6	2.5 M 1,6-Hexanediol
20.	None	0.1 M MES monohydrate pH 6.5	1.6 M Magnesium sulfate heptahydrate
21.	0.1 M Sodium phosphate monobasic monohydrate 0.1 M Potassium phosphate monobasic	0.1 M MES monohydrate pH 6.5	2.0 M Sodium chloride
22.	None	0.1 M MES monohydrate pH 6.5	12% w/v Polyethylene glycol 20,000
23.	1.6 M Ammonium sulfate	0.1 M MES monohydrate pH 6.5	10% v/v 1,4-Dioxane
24.	0.05 M Cesium chloride	0.1 M MES monohydrate pH 6.5	30% v/v Jeffamine® M-600®
25.	0.01 M Cobalt(II) chloride hexahydrate	0.1 M MES monohydrate pH 6.5	1.8 M Ammonium sulfate
26.	0.2 M Ammonium sulfate	0.1 M MES monohydrate pH 6.5	30% w/v Polyethylene glycol monomethyl ether 5,000
27.	0.01 M Zinc sulfate heptahydrate	0.1 M MES monohydrate pH 6.5	25% v/v Polyethylene glycol monomethyl ether 550
28.	None	None	1.6 M Sodium citrate tribasic dihydrate pH 6.5
29.	0.5 M Ammonium sulfate	0.1 M HEPES pH 7.5	30% v/v (+/-)-2-Methyl-2,4-pentanediol
30.	None	0.1 M HEPES pH 7.5	10% w/v Polyethylene glycol 6,000 5% v/v (+/-)-2-Methyl-2,4-pentanediol
31.	None	0.1 M HEPES pH 7.5	20% v/v Jeffamine® M-600®
32.	0.1 M Sodium chloride	0.1 M HEPES pH 7.5	1.6 M Ammonium sulfate
33.	None	0.1 M HEPES pH 7.5	2.0 M Ammonium formate
34.	0.05 M Cadmium sulfate hydrate	0.1 M HEPES pH 7.5	1.0 M Sodium acetate trihydrate
35.	None	0.1 M HEPES pH 7.5	70% v/v (+/-)-2-Methyl-2,4-pentanediol
36.	None	0.1 M HEPES pH 7.5	4.3 M Sodium chloride
37.	None	0.1 M HEPES pH 7.5	10% w/v Polyethylene glycol 8,000 8% v/v Ethylene glycol
38.	None	0.1 M HEPES pH 7.5	20% w/v Polyethylene glycol 10,000
39.	0.2 M Magnesium chloride hexahydrate	0.1 M Tris pH 8.5	3.4 M 1,6-Hexanediol
40.	None	0.1 M Tris pH 8.5	25% v/v tert-Butanol
41.	0.01 M Nickel(II) chloride hexahydrate	0.1 M Tris pH 8.5	1.0 M Lithium sulfate monohydrate
42.	1.5 M Ammonium sulfate	0.1 M Tris pH 8.5	12% v/v Glycerol
43.	0.2 M Ammonium phosphate monobasic	0.1 M Tris pH 8.5	50% v/v (+/-)-2-Methyl-2,4-pentanediol
44.	None	0.1 M Tris pH 8.5	20% v/v Ethanol
45.	0.01 M Nickel(II) chloride hexahydrate	0.1 M Tris pH 8.5	20% w/v Polyethylene glycol monomethyl ether 2,000
46.	0.1 M Sodium chloride	0.1 M BICINE pH 9.0	20% v/v Polyethylene glycol monomethyl ether 550
47.	None	0.1 M BICINE pH 9.0	2.0 M Magnesium chloride hexahydrate
48.	None	0.1 M BICINE pH 9.0	2% v/v 1,4-Dioxane 10% w/v Polyethylene glycol 20,000

Molecular Dimensions

Soluble Proteins

JCSG-*plus* Conditions 1-48 (Box 1) MD1–37

Tube #	Conc.	Salt	Conc.	Buffer	pH	Conc.	Precipitant
1-1	0.2 M	Lithium sulfate	0.1 M	Sodium acetate	4.5	50 % w/v	PEG 400
1-2		None	0.1 M	Sodium citrate	5.5	20 % w/v	PEG 3000
1-3	0.2 M	Ammonium citrate dibasic		None		20 % w/v	PEG 3350
1-4	0.02 M	Calcium chloride dihydrate	0.1 M	Sodium acetate	4.6	30 % v/v	MPD
1-5	0.2 M	Magnesium formate dihydrate		None		20 % w/v	PEG 3350
1-6	0.2 M	Lithium sulfate	0.1 M	Phosphate/citrate	4.2	20 % w/v	PEG 1000
1-7		None	0.1 M	CHES	9.5	20 % w/v	PEG 8000
1-8	0.2 M	Ammonium formate		None		20 % w/v	PEG 3350
1-9	0.2 M	Ammonium chloride		None		20 % w/v	PEG 3350
1-10	0.2 M	Potassium formate		None		20 % w/v	PEG 3350
1-11	0.2 M	Ammonium phosphate monobasic	0.1 M	Tris	8.5	50 % v/v	MPD
1-12	0.2 M	Potassium nitrate		None		20 % w/v	PEG 3350
1-13	0.8 M	Ammonium sulfate	0.1 M	Citrate	4.0		None
1-14	0.2 M	Sodium thiocyanate		None		20 % w/v	PEG 3350
1-15		None	0.1 M	BICINE	9.0	20 % w/v	PEG 6000
1-16		None	0.1 M	HEPES	7.5	10 % w/v	PEG 8000
						8 % v/v	Ethylene glycol
1-17		None	0.1 M	Sodium cacodylate	6.5	40 % v/v	MPD
						5 % w/v	PEG 8000
1-18		None	0.1 M	Phosphate/citrate	4.2	40 % v/v	Ethanol
						5 % w/v	PEG 1000
1-19		None	0.1 M	Sodium acetate	4.6	8 % w/v	PEG 4000
1-20	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	7.0	10 % w/v	PEG 8000
1-21		None	0.1 M	Citrate	5.0	20 % w/v	PEG 6000
1-22	0.2 M	Magnesium chloride hexahydrate	0.1 M	Sodium cacodylate	6.5	50 % v/v	PEG 200
1-23	1.6 M	Sodium citrate tribasic dihydrate pH 6.5		None			None
1-24	0.2 M	Potassium citrate tribasic monohydrate		None		20 % w/v	PEG 3350
1-25	0.2 M	Sodium chloride	0.1 M	Phosphate/citrate	4.2	20 % w/v	PEG 8000
1-26	1.0 M	Lithium chloride	0.1 M	Citrate	4.0	20 % w/v	PEG 6000
1-27	0.2 M	Ammonium nitrate		None		20 % w/v	PEG 3350
1-28		None	0.1 M	HEPES	7.0	10 % w/v	PEG 6000
1-29	0.8 M	Sodium phosphate monobasic monohydrate	0.1 M	Sodium HEPES	7.5		None
	0.80 M	Potassium phosphate monobasic					
1-30		None	0.1 M	Phosphate/citrate	4.2	40 % v/v	PEG 300
1-31	0.2 M	Zinc acetate dihydrate	0.1 M	Sodium acetate	4.5	10 % w/v	PEG 3000
1-32		None	0.1 M	Tris	8.5	20 % v/v	Ethanol
1-33		None	0.1 M	Sodium/potassium phosphate	6.2	25 % v/v	1,2-Propandiol
						10 %v/v	Glycerol
1-34		None	0.1 M	BICINE	9.0	10 % w/v	PEG 20,000
						2 % v/v	1,4-Dioxane
1-35	2.0 M	Ammonium sulfate	0.1 M	Sodium acetate	4.6		None
1-36		None		None		10 % w/v	PEG 1000
						10 % w/v	PEG 8000
1-37		None		None		24 % w/v	PEG 1500
						20 % v/v	Glycerol
1-38	0.2 M	Magnesium chloride hexahydrate	0.1 M	Sodium HEPES	7.5	30 % v/v	PEG 400
1-39	0.2 M	Sodium chloride	0.1 M	Sodium/potassium phosphate	6.2	50 % v/v	PEG 200
1-40	0.2 M	Lithium sulfate	0.1 M	Sodium acetate	4.5	30 % w/v	PEG 8000
1-41		None	0.1 M	HEPES	7.5	70 % v/v	MPD
1-42	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	8.5	20 % w/v	PEG 8000
1-43	0.2 M	Lithium sulfate	0.1 M	Tris	8.5	40 % v/v	PEG 400
1-44		None	0.1 M	Tris	8.0	40 % v/v	MPD
1-45	0.17 M	Ammonium sulfate		None		25.5 % w/v	PEG 4000
						15 % v/v	Glycerol
1-46	0.2 M	Calcium acetate hydrate	0.1 M	Sodium cacodylate	6.5	40 % v/v	PEG 300
1-47	0.14 M	Calcium chloride dihydrate	0.07 M	Sodium acetate	4.6	14 % v/v	2-Propanol
						30 % v/v	Glycerol
1-48	0.04 M	Potassium phosphate monobasic		None		16 % w/v	PEG 8000
						20 % v/v	Glycerol



JCSG-*plus* Conditions 1-48 (Box 2) MD1–37

Tube #	Conc.	Salt	Conc.	Buffer	pH	Conc.	Precipitant
2-1	1.0 M	Sodium citrate tribasic dihydrate	0.1 M	Sodium cacodylate	6.5		None
2-2	2.0 M	Ammonium sulfate	0.1 M	Sodium cacodylate	6.5		None
	0.2 M	Sodium chloride					
2-3	0.2 M	Sodium chloride	0.1 M	HEPES	7.5	10 % v/v	2-Propanol
2-4	1.26 M	Ammonium sulfate	0.1 M	Tris	8.5		None
	0.2 M	Lithium sulfate					
2-5		None	0.1 M	CAPS	10.5	40 % v/v	MPD
2-6	0.2 M	Zince acetate dihydrate	0.1 M	Imidazole	8.0	20 % w/v	PEG 3000
2-7	0.2 M	Zinc acetate dihydrate	0.1 M	Sodium cacodylate	6.5	10 % v/v	2-Propanol
2-8	1.0 M	Ammonium phosphate dibasic	0.1 M	Sodium acetate	4.5		None
2-9	1.6 M	Magneisum sulfate heptahydrate	0.1 M	MES	6.5		None
2-10		None	0.1 M	BICINE	9.0	10 % w/v	PEG 6000
2-11	0.16 M	Calcium acetate hydrate	0.08 M	Sodium cacodylate	6.5	14.4 % w/v	PEG 8000
						20 % v/v	Glycerol
2-12		None	0.1 M	Imidazole	8.0	10 % w/v	PEG 8000
2-13	0.05 M	Cesium chloride	0.1 M	MES	6.5	30 % v/v	Jeffamine® M-600
2-14	3.2 M	Ammonium sulfate	0.1 M	Citrate	5.0		None
2-15		None	0.1 M	Tris	8.0	20 % v/v	MPD
2-16		None	0.1 M	HEPES	7.5	20 % v/v	Jeffamine® M-600
2-17	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	8.5	50 % v/v	Ethylene glycol
2-18		None	0.1 M	BICINE	9.0	10 % v/v	MPD
2-19	0.8 M	Succinic acid pH 7.0		None			None
2-20	2.1 M	DL-Malic acid pH 7.0		None			None
2-21	2.4 M	Sodium malonate dibasic monohydrate pH 7.0		None			None
2-22	1.1 M	Sodium malonate dibasic monohydrate	0.1 M	HEPES	7.0	0.5 % v/v	Jeffamine® ED-2003
2-23	1.0 M	Succinic acid	0.1 M	HEPES	7.0	1 % w/v	PEG 2000 MME
2-24		None	0.1 M	HEPES	7.0	30 % v/v	Jeffamine® M-600
2-25		None	0.1 M	HEPES	7.0	30 % v/v	Jeffamine® ED-2003
2-26	0.02 M	Magnesium chloride hexahydrate	0.1 M	HEPES	7.5	22 % w/v	Poly(acrylic acid sodium salt) 5100
2-27	0.01 M	Cobalt(II) chloride hexahydrate	0.1 M	Tris	8.5	20 % w/v	Polyvinylpyrrolidone
2-28	0.2 M	TMAO	0.1 M	Tris	8.5	20 % w/v	PEG 2000 MME
2-29	0.005 M	Cobalt(II) chloride hexahydrate	0.1 M	HEPES	7.5	12 % w/v	PEG 3350
	0.005 M	Cadmium chloride hemi(pentahydrate)					
	0.005 M	Magnesium chloride hexahydrate					
	0.005 M	Nickel(II) chloride hexahydrate					
2-30	0.2 M	Sodium malonate dibasic monohydrate		None		20 % w/v	PEG 3350
2-31	0.1 M	Succinic acid		None		15 % w/v	PEG 3350
2-32	0.15 M	DL-Malic acid		None		20 % w/v	PEG 3350
2-33	0.1 M	Potassium thiocyanate		None		30 % w/v	PEG 2000 MME
2-34	0.15 M	Potassium bromide		None		30 % w/v	PEG 2000 MME
2-35	2.0 M	Ammonium sulfate	0.1 M	BIS-Tris	5.5		None
2-36	3.0 M	Sodium chloride	0.1 M	BIS-Tris	5.5		None
2-37	0.3 M	Magnesium formate dihydrate	0.1 M	BIS-Tris	5.5		None
2-38	1.0 M	Ammonium sulfate	0.1 M	BIS-Tris	5.5	1 % w/v	PEG 3350
2-39		None	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
2-40	0.2 M	Calcium chloride dihydrate	0.1 M	BIS-Tris	5.5	45 % v/v	MPD
2-41	0.2 M	Ammonium acetate	0.1 M	BIS-Tris	5.5	45 % v/v	MPD
2-42	0.1 M	Ammonium acetate	0.1 M	BIS-Tris	5.5	17 % w/v	PEG 10,000
2-43	0.2 M	Ammonium sulfate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
2-44	0.2 M	Sodium chloride	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
2-45	0.2 M	Lithium sulfate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
2-46	0.2 M	Ammonium acetate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
2-47	0.2 M	Magnesium chloride hexahydrate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
2-48	0.2 M	Ammonium acetate	0.1 M	HEPES	7.5	45 % v/v	MPD