FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

ICBAS | INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
SCHOOL OF MEDICINE AND BIOMEDICAL SCIENCES

MSc

2.º
CICLO

FCUP
ICBAS
2021

U.PORTO

Understanding Structure-Function Correlations of Enzymes with Biotechnological Relevance

Inês Oliveira Justo

FC

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

ICBAS | INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
SCHOOL OF MEDICINE AND BIOMEDICAL SCIENCES

Understanding Structure-Function Correlations of Enzymes with Biotechnological Relevance

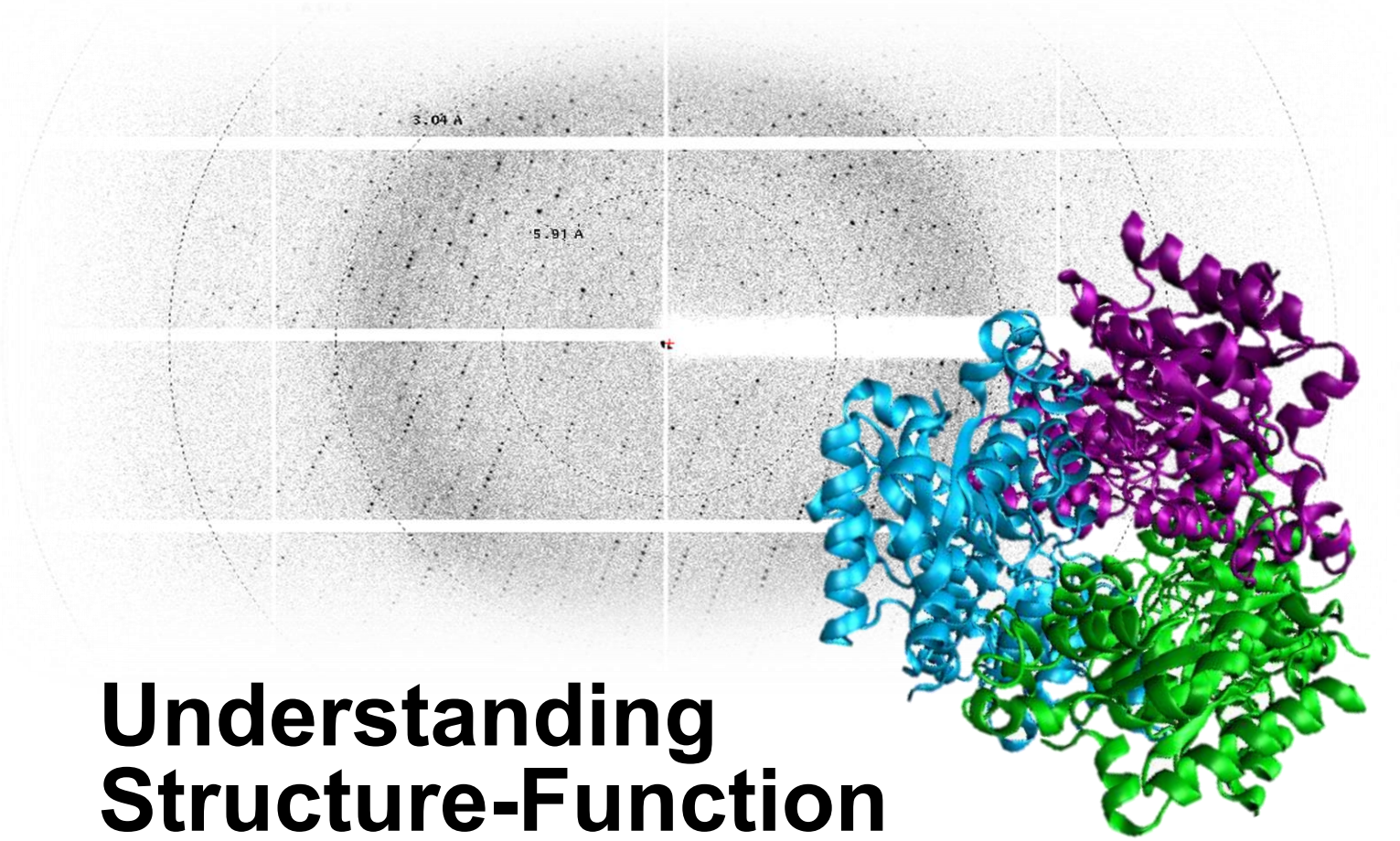
Inês Oliveira Justo

Master thesis presented to Faculty of Sciences (FCUP) and Abel Salazar's Institute of Biomedical Sciences (ICBAS), University of Porto

MSc in Biochemistry

2021

MSc



Understanding Structure-Function Correlations of Enzymes with Biotechnological Relevance

Inês Oliveira Justo

MSc in Biochemistry

FCUP/ICBAS, Porto University

European Molecular Biology Laboratory (EMBL), Hamburg

2021

Supervisor
Dr. Isabel Bento, Staff Scientist, EMBL-Hamburg

Co-supervisor
Dr. Luís Gales Pinto, Associate Professor, ICBAS

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Acknowledgements

Surprisingly, writing this seems to be the hardest part. My individual success is only a testament of all the little contributions of everyone in my life and putting that enormity of a sentiment into words seems almost like a disservice. However, like with most things, I will try.

First and foremost, I would like to thank my supervisor and co-supervisor, Dr. Isabel Bento and Prof. Luís Gales, for giving me the opportunity to do my master thesis under their guidance. Working at EMBL-Hamburg has been one of the great joys of my life and part of it was the ability to work directly with Isabel. Thank you for your guidance and encouragement to make me a better person and a better scientist. Thank you for all your patience and for never once undermining my thoughts and ideas. Most of all, thank you for your friendship and your mentorship, I will never forget it.

A special thanks to everyone at EMBL-Hamburg for making this journey such a pleasure. Specially my group (Schneider Group) and Garcia Group, which “adopted” me most socially distanced lunchtimes. My stay in Hamburg was definitely made better by the presence of Janina, George, Christian, Lucas, Javi, Katha, Stephan, Angelica and Lola. Thank you all.

To my friends in Portugal, that never stopped believing in me. From Lisbon I bring with me friends that are more like family: Bárbara Costa, Ana Catarina Oliveira, Andreia Barro, Marta Rosa, Rafael Vieira, Dinis Rodrigues e Francisco Cascalho. In Porto I met wonderful people that always encouraged me to be and to do better: Alexandra Borges, Juliana Amorim, Alexandra Simões, Ana Carvalho, Ana Moura, João Varela, Beatriz Lima and Joana Castro. There are many more but I could not possibly list everyone. To all “nortenhos” e “FCTenses”, I’ll bring you with me wherever I go next.

Last but definitely not least, thank you to my brother and father who don’t quite understand what I do or study but try to be supportive of it just the same. And most importantly, thank you to my Grandma (Vó) for being the best mother anyone could ask for and for calling EVERY DAY to check if I’m feeding myself right. Who knows where I would be without you.

Without you none of this would’ve been possible.

Obrigada a todos.

Abstract

The basis of all organic compounds are the carbon-carbon bonds that form them. In this regard, enzymatic aldol reactions offer a great alternative to the industrial synthesis of organic compounds, which is often extremely hard, pollutant and expensive. Aldolases are enzymes capable of catalyzing these carbon-carbon bond formation reactions, in a reversible and stereospecific manner, making them a great tool for the synthesis of organic molecules at the industrial level. Therefore, the biochemical, biophysical and structural study of these enzymes could be very beneficial for different industries (e.g., pharmaceutical and food industries).

We focus on class II pyruvate aldolases, which use a divalent metal cation as a cofactor and pyruvate as the nucleophile donor substrate. Although usually donor-specific, it has been shown that besides pyruvate, some of these enzymes can also use hydroxypyruvate (HPA) as a donor substrate. Structural studies were conducted with two different stereo-complementary bacterial class II pyruvate aldolases: HpcH/HpaI aldolase from Sphingomonas wittichii RW1 (*Sw*HKA) and 2-dehydro-3-deoxyglucarate aldolase from Burkholderia phytofirmans (*Bp*HKA), which exclusively accept and form D and L-enantiomers, respectively. The goal of this project was to try to understand the structural determinants that allow for this specific selectivity regarding the acceptor aldehyde substrate.

This study involved the use of techniques such as SDS-PAGE, differential scanning fluorimetry (nDSF), the screening for optimal crystallization conditions and X-ray structural determination of the apo and holo forms of *Sw*HKA and *Bp*HKA in complex with HPA, as well as SwHKA mutants. After the refinement of the structures, the results obtained revealed the importance of a serine residue positioned near the active site of *Sw*HKA, allowing the isolation and characterization of a new conformation state of this enzyme.

Keywords: class II pyruvate aldolase, X-ray crystallography, crystal structure determination, magnesium cofactor, resting state conformation, HpcH/HpaI aldolase, 2-dehydro-3-deoxyglucarate aldolase.

Resumo

A base de todos os compostos orgânicos são as ligações carbono-carbono que os formam. Assim, reações enzimáticas de aldol são uma ótima alternativa à síntese destes compostos a nível industrial, normalmente bastante difícil, cara e poluente. As aldolases são enzimas que catalisam reações de síntese de ligações carbono-carbono, de uma maneira reversível e estéreo-específica, tornando-as uma ótima alternativa para a síntese de compostos orgânicos a nível industrial. Assim, o estudo bioquímico, biofísico e estrutural destas enzimas pode ser bastante benéfico para diferentes indústrias (ex. indústria farmacêutica e alimentar).

Focámo-nos em aldolases de piruvato de classe II, que usam catiões metálicos divalentes como cofator e piruvato como substrato dador nucleofílico. Apesar de normalmente serem bastante específicas quanto ao dador, já foi estudado que para além de piruvato algumas destas enzimas também conseguem usar hidroxipiruvato (HPA) como substrato dador. Neste projeto foram feitos estudos estruturais com duas aldolases de piruvato de classe II estéreo-complementares: aldolase HpcH/HpaI da bactéria *Sphingomonas wittichii* RW1 (*Sw*HKA) e a aldolase 2-dehidro-3-deoxiglucarato da bacteria *Burkholderia phytofirmans* (*Bp*HKA), que aceitam/formam exclusivamente enantiómeros D e L, respetivamente. Este trabalho teve como objetivo tentar perceber os elementos estruturais que permitem esta promiscuidade relativamente ao substrato aceitador (aldeído).

Este estudo envolveu o uso de técnicas como SDS-PAGE, scan de fluorimétrica diferencial (nDSF), procura de condições de cristalização ótimas e determinação estrutural de raios-X dos estados *apo* e *holo* da *Sw*HKA e *Bp*HKA em complexo com HPA, tal como mutantes da *Sw*HKA. Após o refinamento das estruturas, os resultados obtidos revelaram a importância de um resíduo de serina posicionado perto do centro ativo da *Sw*HKA, permitindo o isolamento e caracterização de um novo estado conformacional desta enzima.

Palavras-chave: aldolase de piruvato de classe II, cristalografia de raio-X, determinação estrutural cristalina, cofator de magnésio, conformação de estado de repouso, aldolase HpcH/HpaI, aldolase 2-dehidro-3-deoxiglucarato.

Table of Content

Acknowledgements I

Abstract II

Resumo III

Table of Content IV

List of Figures VI

List of Tables IX

List of f Abbreviations, symbols and acronyms X

1. Introduction 1

1.1. Macromolecule crystallography 1

1.1.1. Crystallization 2

1.1.2. Crystallization methods 4

1.1.3. Crystallization screening 5

1.1.4. Protein-ligand complexes 6

1.1.5. Seeding 7

1.1.6. Cryoprotection 7

1.1.7. Data collection and structure determination 9

1.2. Class II pyruvate aldolases 13

2. Methods and materials 16

2.1. General procedures 16

2.2. Quality control: SDS-PAGE 16

2.3. Thermal Stability: nDSF assays 17

2.4. Crystallizations screens and trials 21

2.5. Final crystallization and soaking condition 26

2.5.1. *Sw*HKA 26

2.5.2. *Bp*HKA 27

2.6. X-ray data collection 28
2.7. X-ray data processing 28
2.8. Structure solution 28
2.9. Other techniques 29
3. Results and discussion 30
3.1. Quality Control: SDS-PAGE 30
3.1.1. *Sw*HKA 30
3.1.2. *Bp*HKA 31
3.2. Thermal stability: nDSF assays 34
3.2.1. *Sw*HKA 34
3.2.2. *Bp*HKA 37
3.3. Crystallization screens and trials 40
3.3.1. *Sw*HKA 40
3.3.2. *Bp*HKA 43
3.4. Overall crystal structures 52
3.4.1. *Sw*HKA 52
3.4.2. *Bp*HKA 63
3.4.3. Comparison between SwHKA and BpHKA 65
4. Conclusion 70
5. References 72
6. Supplementary data 76

List of Figures

Figure 1 - Schematic illustration of a protein crystallization phase diagram. The four major crystallization methods are represented: (i) microbatch, (ii) vapor diffusion, (iii) dialysis and (iv) free interface diffusion (FID). This will be explained further on. The filled black circles represent the starting conditions. This figure was taken from Figure 2 of reference 4. 3
Figure 2 - Drops of SwHKA-S116A mutant under different crystallization conditions. A - undersaturation; B - crystals; C - nucleation; D - precipitant. 3
Figure 3 - a) The diffraction experiment. b) Equation of Fourier's summation. V is the volume of the unit cell and αhkl is the phase associated with the structure-factor amplitude |Fhkl|. Adapted from Figure 1 of reference 13. 10
Figure 4 - Representation of the general structure of class II pyruvate aldolases (PDB ID 6R62). 14
Figure 5 - Representation of the active site location of class II pyruvate aldolases (PDB ID 6R62). 14
Figure 6 - Reaction mechanism of SwHKA and BpHKA with hydroxypyruvate as the nucleophilic donor substrate. Adapted from reference 30. 15
Figure 7 - Typical differential scanning fluorimetry (DSF) assay by thermal denaturation. Figure adapted from a BioRender template. 17
Figure 8 - SDS-PAGE gel of S116A and S116C mutants of SwHKA. From left to right: protein ladder, S116A protein sample, S116C protein sample. 30
Figure 9 - SDS-PAGE gel of BpHKA. The samples used were stored for different periods of time at 4ºC and at -80ºC. Time stored at 4ºC. From left to right: approximately 1 year, 1 month, 1 week and 0h (sample defrosted directly before the gel was set-up). 31
Figure 10 - SDS-PAGE gel of BpHKA incubated with and without $MgCl_2$. 33
Figure 11 - Crystallizations hits of S116A and S116C mutants of SwHKA, from position G4 of the Classic I Suit screen. 40
Figure 12 - Drops of SwHKA-S116A (A to D) and S116C-SwHKA (E and F). Drop conditions: A and E - 0.2M $MgCl_2$, 0.1M HEPES (pH 7), 25% PEG400; B - 0.1M $MgCl_2$, 0.1M HEPES pH 7.5, 20% PEG400; C and F - 0.1M $MgCl_2$, 0.1M HEPES pH 8, 20% PEG400. D – crystal close-up of C. 41
Figure 13 - Drop of SwHKA-WT in citrate conditions. 42
Figure 14 - Drops of BpHKA with PEG1500 as the precipitant. Reservoir solution (RS) also consisted of 0.1M bis-tris propane at pH 8.5 and 0.1M $MgCl_2$, at 19ºC (A to D) and 4ºC (E and F).

Drop ratio of 1.8µL:1.2µL (P:RS) with protein solution at 20.8mg/mL (A), 10mg/mL (B to E) and 8mg/mL (F). Percentage (v/v) of PEG1500: A and B– 20%; C – 18%; D to F – 16%..... 43
Figure 15 - Drop of BpHKA with PEG3350 as the precipitant. Reservoir solution (RS) also consisted of 0.2M $MgCl_2$, at 19ºC. Drop ratio of 1.5µL:1.5µL (P:RS) with protein solution at 8mg/mL (A to C), 10mg/mL (D). Percentage (v/v) of PEG3350: A– 14%; B – 12%; C and D – 10%. 45
Figure 16 - Drop of BpHKA with 8% PEG3350 and 0.2M $MgCl_2$, at 19ºC. Drop ratio of 1.5µL:1.5µL (P:RS) with protein solution at 8mg/mL. 46
Figure 17 - Drop of BpHKA with different PEGs as the precipitant. Drop ratio of 1.5µL:1.5µL (P:RS), at 19ºC, with protein solution at 9.5mg/mL. Reservoir solution (RS): A - 25% PEG500MME and 0.1$MgCl_2$; B - 25% PEG500MME and 0.2$MgCl_2$; C - 8% PEG2000MME and 0.1$MgCl_2$; D - 8% PEG4000 and 0.1$MgCl_2$; E - 8% PEG5000MME and 0.1$MgCl_2$..... 47
Figure 18 - Drop of BpHKA with seeding. Reservoir solution (RS) with 8% PEG2000MME (A and B) and 8% PEG4000 (C and D), and 0.1M $MgCl_2$ (all). Drop ratio of 1.5µL:1.5µL (P:RS), at 19ºC, with protein solution at 8mg/mL (A and C) and 7mg/mL (B and D). A and C – no seeding; B and D: seeding with seeds from the crystals in Figure 17. 48
Figure 19 - Drop of BpHKA set with the “Chinese” drop technique. Reservoir solution (RS) with 8% PEG2000MME and 0.1M $MgCl_2$, at 19ºC. Drop ratio of 1.5µL:1.5µL (P:RS) with protein solution at 8mg/mL..... 49
Figure 20 - Crystallizations hits of BpHKA, from position H5, H7 and F9 of the Additive screen. UV inspection confirms the presence of protein crystals. Additive: A - 4% formamide; B - 0.5% ethyl acetate; C - 2% benzamidine hydrochloride. 50
Figure 21 - Drop of BpHKA after a forced dehydration experiment. All drops were set with reservoir solution (RS) of 8% PEG2000MME and 0.1M $MgCl_2$, at 19ºC, drop ratio of 1.5µL:1.5µL (P:RS) and protein solution at 8mg/mL. New PEG2000MME concentration: A – control, no well change; B – 12%; C – 16%; D – 20%. 51
Figure 22 - Drop of BpHKA with 8% PEG5000MME and 0.2M $MgCl_2$, at 19ºC. Drop ratio of 1.5µL:1.5µL (P:RS) with protein solution at 9.5mg/mL. 52
Figure 23 - Schematic overview of the refined SwHKA structures. HPA – hydroxypyruvate; 3GR – D-glyceraldehyde; 8YV – 2-keto-D-gluconic acid (product of the aldol reaction catalyzed by SwHKA). 55
Figure 24 - Representation of the active site of SwHKA-WT, from different structures: amino acids S116, E145 and D171, water molecules (w), magnesium ion (M), hydroxypyruvate (HPA) and 2-keto-D-gluconic acid (8YV). Magnesium, HPA and 8YV are present in the structures at full

occupancy. A - superimposed structures of apo-SwHKA (yellow), SwHKA-Mg (pink), SwHKA-Mg-HPA (cyan) and SwHKA-all (green). B - zoom-in of A, without HPA and 8YV representation. C - apo-SwHKA; D - SwHKA-Mg; E - SwHKA-Mg-HPA; F - SwHKA-all. 56
Figure 25 - Superimposed representation of the active site of different structures of SwHKA-WT. Water molecules not shown for schematic clarity. A - apo-SwHKA (yellow) and SwHKA-Mg (pink); B – SwHKA-Mg (pink) and SwHKA-Mg-HPA (cyan); C - SwHKA-Mg (pink), SwHKA-Mg-HPA (cyan) and SwHKA-all (green). 58
Figure 26 - Representation of the active site of SwHKA-S116A-Mg. 59
Figure 27 - Superimposed representation of the active sites of SwHKA-Mg (pink) and S116A-Mg (blue). Water molecules not shown. A – chain A active site of S116A-Mg; B – chain B active site of S116A-Mg. 60
Figure 28 - Representation of the active site of SwHKA-S116A-Mg-HPA..... 61
Figure 29 - Superimposed representation of the active sites of SwHKA-Mg-HPA (cyan) and S116A-Mg-HPA (orange). Water molecules not shown. A – chain A active site of S116A-Mg-HPA; B – chain B active site of S116A-Mg-HPA. 62
Figure 30 - Superimposed representation of the active sites of S116A-Mg (blue) and S116A-Mg-HPA (orange). Water molecules not shown. A – chain A active site of both structures; B – chain B active site of both structures. 63
Figure 31 - Representation of the active sites of BpHKA-Mg. 64
Figure 32 - Superimposed representation of the equivalent metal (M) coordination active sites of BpHKA-Mg (green), SwHKA-Mg (pink) and S116A-Mg (blue). Water molecules (w) not shown in panel D for schematic clarity. A – SwHKA-Mg; B – Chain B of S116A-Mg; C – Chain A of BpHKA-Mg; D – BpHKA-Mg, SwHKA-Mg and S116A-Mg. 66

List of Tables

Table 1 - Custom nDSF assays of SwHKA. Assays done in triplicate. The magnesium ion (Mg) was added in the form of $MgCl_2$. The samples were incubated for 45 minutes before measurement. 19

Table 2 - Custom nDSF assays of BpHKA. Assays done in triplicate. The magnesium (Mg) and sodium (Na) ions added in the form of $MgCl_2$ and NaCl, respectively. The samples were incubated for 45 minutes before measurement, unless stated otherwise. 20

Table 3 - Commercially available screens used. In the BpHKA screens the base reservoir solution was composed of 8% PEG2000-MME and 0.1M $MgCl_2$. 21

Table 4 - Crystallization conditions tried for SwHKA-S116A and SwHKA-S116C. Total of 108 manually set drops. 22

Table 5 - Crystallization conditions of BpHKA. Total of 326 drops. All drops were manually set except the conditions with a protein(P):reservoir solution(RS) ratio of 1:1, which were set with Mosquito®LCP. Drops marked with an asterisk (*) represent conditions where both "chinese" and "normal" drop set-ups were used. 23

Table 6 - Melting temperatures (T_m) obtained through nDSF for the S116A and S116C mutants of SwHKA, as well as SwHKA-WT. 34

Table 7 - Melting temperatures (T_m) obtained through nDSF for SwHKA-WT when incubated with 0 to 500mM of $MgCl_2$, 0 to 275mM of HPA and 3 different aldehydes. 35

Table 8 - Melting temperatures (T_m) obtained through nDSF for BpHKA-WT. Magnesium (Mg) and sodium (Na) ions were added in the form of $MgCl_2$ and NaCl, respectively. 37

Table 9 - Melting temperatures (T_m) obtained through nDSF for BpHKA-WT when incubated with 0 to 1.9M of $MgCl_2$, 0 to 275mM of HPA and 3 different aldehydes. 38

Table 10 - Data collection and refinement statistics of SwHKA-WT, structures apo-SwHKA and SwHKA-all, and SwHKA-S116A, structures S116A-Mg and S116A-Mg-HPA. Parameters with an asterisk “*” correspond to outer shell values. 53

Table 11 - Data collection statistics of BpHKA-Mg. The refinement process of this structure is not complete so no refinement statistics are shown. Parameters in parenthesis “()” correspond to outer shell values. 64

Table 12 - RMSD and sequence identity values between the protein chains with equivalent metal (M) coordination active sites of SwHKA-Mg (chain A), S116A-Mg (chain B) and BpHKA-Mg (chain A). 67

List of f Abbreviations, symbols and acronyms

***Sw*HKA**	HpcH/HpaI aldolase from *Sphingomonas wittichii* RW1
***Bp*HKA**	2-dehydro-3-deoxyglucarate aldolase from *Burkholderia phytofirmans*
3D	Three dimensional
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
PDB	Protein Data Bank
FID	Free interface diffusion
L	Litre
mL	Millilitre
µL	Microlitre
nL	Nanolitre
K	Kelvin degrees
ºC	Celsius degrees
PEG	Polietilenoglicol
$\|F_{hkl}\|$	Amplitude of the wave
V	Volume of the unit cell
α_{hkl}	Phase of the wave
MR	Molecular resolution
RMSD	Root-mean-square deviation
Å	Ångström
MIR	Multiple isomorphous replacement
SAD	Single-wavelength anomalous diffraction
MAD	Multiple-wavelength anomalous diffraction
XFEL	X-ray free electron lasers
μm^3	Cubic micrometer
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
DSF	Differential scanning fluorimetry
KDa	Kilodalton
SPC	Sample and preparation characterization lab
mM	Millimolar
M	Molar

ºC	Celsius degrees
mg	milligram
RCF	Relative centrifugal force
g	g-force
mm	millimeter
nm	nanometer
MES	2-(N-morpholino)ethanesulfonic acid
TSA	Thermal shift assay
T_m	Melting temperature
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
v/v	Volume/volume
w/v	Weight/volume
MW	Molecular weight
WT	Wild-type
P:RS	Protein – reservoir solution drop ratio
$MgCl_2$	Magnesium chloride
HPA	Hydroxypyruvate
8YV	2-keto-D-gluconic acid

1. Introduction

1.1. Macromolecule crystallography

Macromolecule crystallography allows the determination of the 3D structure of different macromolecules like proteins, DNA, RNA and protein-ligand complexes. This technique consists of seven main steps: macromolecule crystallization, X-ray diffraction data collection, solution of the phase problem, refinement of the 3D structure, validation of the structural model and structural analyses (correlating structure with function).

Although an essential way to determine a protein's structure, with more than 158130 structures in the Protein Data Bank (PDB) discovered this way, macromolecular crystallography is not a "crowded" field. This likely comes from the fact that the technical and mathematical aspects of the data treatment may not be completely intuitive/easy to understand by scientists that lack a strong background in macromolecular structure.[1] It can also be a frustrating field at times since there's no correlation between how good a crystal looks and its ability to diffract.[1]

Once an "elite" technique, used only by highly trained scientists, crystallography has become much more accessible due to the development of new equipment and sophisticated computer softwares that provide an easier way to collect and refine diffraction data.[1,2] Besides obtaining and processing the diffraction data, finding optimal crystallization conditions for each macromolecule has also been largely facilitated by crystallization robots like Mosquito® and commercially available screening plates that allow a much faster way to test a lot of different conditions.[1,3]

Today, the large majority of the macromolecules of interest are produced through genetic engineering of expression vectors in bacteria, usually *Escherichia coli*. This contrasts with the past, where macromolecules were isolated directly from the biological source.[1] This new method has the advantage of yielding bigger amounts of protein and allowing the expression of recombinant versions of the desired sequence, making it easier to work with.[4] For example, we can manipulate the original nature-existing sequences in our favor by removing unordered regions or by introducing His-Tag sequences, making the purification process easier.[5] These changes in the amino acid sequence of a given protein (primary structure) will of course alter its tertiary structure, which can also be used to our advantage.

Besides proteins/nucleic acids, it is also possible to crystallize small organic molecules (low molecular mass compounds), but the crystals obtained show different characteristics. On one hand, when compared to small molecules, protein crystals tend to be smaller, disintegrate/crack easily if dehydrated and diffract X-rays poorly. On the other hand, small molecules tend to yield stronger crystals that can usually be exposed to air and have firmer lattice forces that diffract X-rays intensely.[3]

Macromolecules are also harder to crystalize than small molecules, not only because they have different kinetically favored solid states (crystals, precipitant, oils and gels), but also because they tend to be bigger in size, which makes them slower to reach the high levels of supersaturation needed for crystallization.[3] Macromolecule crystals are also sensitive to temperature changes and suffer extensive damage after exposure to radiation.

1.1.1. Crystallization

The crystallization process consists of a controlled phase transition process that generates a 3D structure of periodically repetitive and highly identical motifs (unit cells).[1,5]

This occurs in two steps, nucleation and growth, that require the manipulation of the protein's solubility, forcing a supersaturation state.[3,5] This can be achieved by increasing the ionic strength (adding salts), changing the pH value (causing a change in the protonation state of the amino acid residues), changing the temperature, adding a precipitant agent or a ligand that stabilizes the protein, increasing the protein's concentration or adding a cross-biding agent that promotes lattice interactions.[3,6]

In the nucleation step (either homogeneous or heterogeneous), the molecules in solution go from disordered to an ordered state, seemingly by the formation of a partially ordered intermediate.[3,6] This process yields "critical nuclei", small but ordered aggregates. Then comes the growth step where ideally, as the name says, the critical nuclei spontaneously grow to full-size crystals.[3]

Although separate, both the nucleation and growth steps dependent of the creation of the supersaturated state, making it essential to the crystallization process.[3]

With all proteins in an aqueous solution, no matter the composition, a solubility curve can be established, separating the conditions in which the solution is either undersaturated or supersaturated (Figure 1). Over the solubility curve, we reach a nonequilibrium state where the

amount of protein in solution exceeds the solubility limit but, in certain chemical and physical conditions, is still in solution.[3] In order to re-establish the equilibrium and lowed the oversaturation of the system, a solid state is formed.[3]

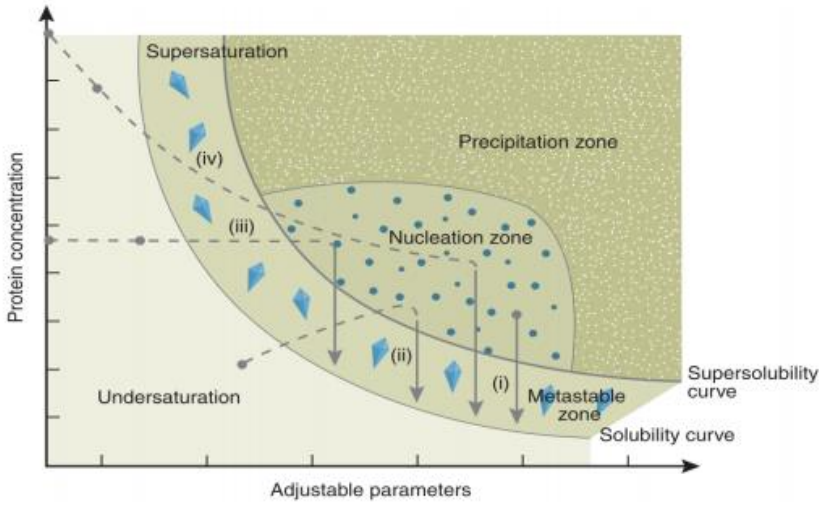


Figure 1 - *Schematic illustration of a protein crystallization phase diagram. The four major crystallization methods are represented: (i) microbatch, (ii) vapor diffusion, (iii) dialysis and (iv) free interface diffusion (FID). This will be explained further on. The filled black circles represent the starting conditions. This figure was taken from Figure 2 of reference 4.*

As we can see in Figure 1, the supersaturation state can be divided in 3 different zones: a high supersaturation zone where the protein precipitates (precipitation zone), a zone of medium supersaturation where the nucleation step occurs (nucleation zone) and a lower supersaturation zone where the nucleation stops and the growth step begins spontaneously (metastable zone).[5,6] These different states and zones can be easily observed under the microscope since each one of them has a different appearance (Figure 2): clear drop (undersaturation), crystals (metastable zone), crystalline precipitant (nucleation zone) and amorphous precipitant (precipitation zone). These visible differences allow the easy monitorization of the crystallization process after drop set-up.

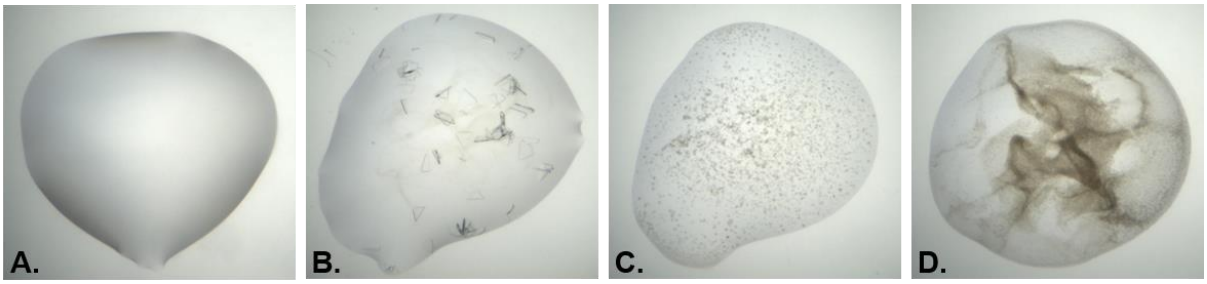


Figure 2 - *Drops of SwHKA-S116A mutant under different crystallization conditions. A - undersaturation; B - crystals; C - nucleation; D - precipitant.*

1.1.2. Crystallization methods

Depending on the way the kinetics of the phase diagram are manipulated, there are a few different crystallization methods possible.[6] The most common ones are vapor diffusion, microbatch experiments, dialysis and free interface diffusion (FID).

Vapor diffusion is, and has always been, the most used method.[5] In the beginning, when everything was done manually, vapor diffusion was the method that allowed the use of the least amount of protein. Now, along with microbatch, it's one of the easiest methods to automate.

In an isolated environment, a drop (either sitting or hanging) of protein solution and crystallization solution (also called reservoir solution or mother liquid), usually in a ratio of 1:1, equilibrates overtime by vapor diffusion with the crystallization solution in the reservoir, which is at a higher or lower concentration than the one in the drop.[3,5] Although the most used method, vapor diffusion has some disadvantages, like changes in pH value due to volatile components and the uncontrollable change in drop volume/composition during the crystallization process.[5] This makes some conditions unreproducible from drop to drop, especially with the use of volatile components.

In microbatch experiments, the drop conditions stay constant within the time frame of the experiment because the drop (protein plus crystallization solution) is incubated under a low-density paraffin oil (0.87 g/ml) that isolates the drop from contaminants and evaporation.[5] This method allows the achievement of supersaturation as soon as the two solutions are mixed, as opposite of what happens with the other methods where the protein starts off in the undersaturated state (Figure 1).[5]

For dialysis, commercially available dialysis buttons "trap" the protein behind a semipermeable membrane. The protein can't go through the membrane but the crystallization solution can, slowly mixing with the protein overtime.[3,5]

With FID, the protein is layered on top of the crystallization solution, in a capillary with sealed ends. Overtime, a concentration gradient is established between both solutions, inducing nucleation and crystal growth.[3,5] This method is better used as a fine-tuning technique once the best crystal-growing conditions are established.[5]

After large enough crystals are grown, they are harvested using special loops (transparent under the X-rays), soaked in a cryoprotectant solution and frozen in liquid nitrogen. When it's time

to perform the diffraction experiment, the loops with the crystals are mounted and exposed to the X-rays.[1]

1.1.3. Crystallization screening

Finding crystallization conditions that yield good diffraction protein crystals is usually done in two parts: first, different conditions (precipitants, pH values, temperature…) are screened in order to find the ones that give initial crystalline material.[3,6] After that, once even the smallest crystal is observed, it's "only" a matter of fine-tuning those conditions until the best crystal samples for X-ray diffraction are achieved.[3,5] The fine-tuning process consists of gradually making small, systematic changes in the crystallization parameters as an attempt to isolate the exact conditions that yield the best crystals.[3,4]

Crystallization is in essence a trial and error process, since every protein is unique and even proteins from the same family can have dramatically different crystallization behaviors.[3] For this reason, the most common outcome is the formation of clusters of tiny crystals, caused by excess nucleation, or no crystals at all.[6] These results are very far from what it's needed to obtain good diffraction data.

The automation of the screening process is not only possible but also more "economic", using much smaller quantities of protein per drop, in the scale of the nanolitre (10^{-9}L).[3] Since proteins are expensive and hard to produce, the use of less protein sample is an obvious advantage, when compared with the traditional methods. However, crystallization depends on both thermodynamic and kinetic factors and reducing the volumes to such a small scale can sometimes lead to "false negatives" – crystallization doesn't happen in conditions where it usually would if bigger volumes were used.[4,5]

Because of the small scale of this process, the resulting crystals are usually too small to harvest manually. The solution is to set crystal drops with the same conditions and bigger volumes (in the µL scale as opposed to the nL scale). Consequently, this will produce bigger crystals if the right crystallization conditions are used.

1.1.4. Protein-ligand complexes

Understanding the binding mechanisms of a protein with different ligands could be key to designing new drug-targets or understanding different enzymatic mechanisms, making the study of protein-ligand complexes of high interest for both Industry and Academia. However, the crystallization and diffraction analysis of these complexes can be quite difficult since each protein presents a different and unique case. Depending on the protein of interest, different methods can be used to obtain the complex's structure.

Co-expressing the protein with the ligand often results in increased levels of protein expression and amount of soluble protein.[7] Using this method is not always possible since it depends a lot on the solubility and potency of the ligand, as well as its availability.[7]

For optimized expression systems with difficult purification processes, one option to improve the protein's stability and solubility, in some cases, is to add the ligand in the purification process, either in the cell-lysis step or at the refolding stage.[7] However, the most common ways to obtain protein complexes is either by soaking or by co-crystallization.

In co-crystallization, the ligand solution is added to the protein sample at the beginning, either through incubation or by adding the ligand when the crystallization drops are being setup.[7,8] When the crystallization and complex binding are successful, every crystal grown is of the protein-ligand complex.[7,8] This depends of the protein and ligand concentrations, protein-ligand ratio and temperature.[7,8]

With soaking, protein-only crystals are grown and later transferred to a solution with the desired ligand.[8] This method has the advantage of allowing the test of different ligands at the same time because a single drop (in optimized conditions) usually yields more than one crystal and each one of them can be soaked with a different ligand, which decreases the number of crystallization experiments needed.[8]

In some cases, the proteins can crystallize with one ligand but not with others so a "replacement soaking" can be performed, where the protein is crystallized with an initial first ligand and then those crystals are soaked with a second ligand, as long as the protein has more affinity for the second ligand than to the first.[7] Some crystals may be sensitive to solvents used to dissolve the ligands, so in both soaking and co-crystallization, there may be the need to use stabilizing additives.[7,8]

1.1.5. Seeding

One of the hardest parts of the crystallization process is to pinpoint the exact conditions that avoid excess nucleation and yield large, single crystals instead. As stated before, excessive nucleation and the formation of crystalline precipitant (microcrystals) is the most common result of crystallization setups, since it's very difficult to predict the conditions that lead to the metastable zone, allowing crystal growth. When a condition yields microcrystals instead of the desired-sized ones, seeding experiments can be used to separate nucleation from growth.

This technique consists of transferring nuclei (seeds) from one drop, with excessive nucleation, to another with lower supersaturation conditions, in order to promote the growth of the already formed nuclei and avoid further nucleation.[9] The lower supersaturation levels of the second drop are achieved by decreasing the concentration of one of the components, either the protein itself or the precipitating agent.

There are two main types of seeding, depending on the sizes of the seeds: microseeding, where the seeds are too small to separate individually, and macroseeding, where single crystals are yielded but they are too small for data collection.[9]

The most used microseeding method consists of smashing the nuclei/small crystals that grow at first and transferring them to a new drop, where the nucleation process will be skipped and the crystals will grow bigger. The easiest transferring method however is called streak seeding and uses a seeding wand or an animal whisker/hair (cat/horse) to grab the seeds from the original drop.[9] The wand/whisker is then passed through the new drop, depositing the seeds in a streak line. The use of animal whiskers is advantageous due to its overlapping cuticles (smooth surface) that easily capture and release the seeds in the drop solutions.[9]

1.1.6. Cryoprotection

Macromolecules' crystals have limited diffraction pattern resolutions due to the solvent channels that fill the spaces between the molecules and weaken the lattice force of the crystal.

In a protein's case, this large spaces between molecules can allow the slight variations in position of the residues' side chains of the protein within the unit cell.[3,10] This slight change in orientation, when applied to all unit cells, can create polymorphism within the crystal and lowers the overall quality of the diffraction pattern.[3] More so, in their biological systems, proteins operate in aqueous means so the high solvent content of the crystals maintain the protein's structure in

crystal form as if it was in solution, allowing the study of the “normal” interaction between the protein and different ligands, ions, inhibitors and substrates.[3] However, the aqueous content of the solvent channels becomes a problem when the crystals are exposed to X-rays.

At room temperature, when exposed to the ionizing high energy of the X-ray beams, crystals suffer radiation damage, forming free radicals that will move through the solvent channels, damaging the rest of the crystal arrangement and resulting in progressively worse diffraction data quality.[11,12]

At cryogenic temperatures (around 100K) however, the content of the solvent channels is frozen, vastly reducing the diffusion of the free radicals.[11] The damage to the structure is confined to the exact place where the crystal was exposed to radiation and the rest of the crystalline arrangement stays intact. This way, since the overall structure is not affected, different data sets can be acquired from a single crystal just by aiming the beam at different parts of the crystal. This way, less crystals are needed, which saves time and resources.

With the need to collect the diffraction data at cryogenic temperatures surges the need to change the aqueous composition of the solvent channels. At cryo temperatures, water freezes and consequently expands, disturbing the crystalline arrangement and the lattice forces that keep the crystal together.[10] More so, the frozen water molecules will yield a characteristic circular diffraction pattern (ice rings), increasing the background of the collected data and decreasing its quality and accuracy.[10,12]

Thereby, cryoprotection is achieved not only by collecting the diffraction data at low temperatures (under nitrogen flow), but also by using a cryoprotectant. Glycerol and PEG400 are examples of common cryoprotectants that can be used by soaking the crystals in a drop with both the crystallization solution and the cryoprotectant agent.[12] The soaked crystal is then mounted directly on a loop made of X-ray transparent material, submerged in liquid nitrogen and stored. The possibility of storing the crystals for long periods of time prior to data collection, without loss of crystalline arrangement, is also a big advantage.

One disadvantage of this process is that the cryoprotectant conditions for each protein crystal can be as specific as the crystallization conditions and cryoprotectant trials should be made as using the wrong cryoprotectant can destroy the crystalline structure.[11]

1.1.7. Data collection and structure determination

X-rays are a form of electromagnetic radiation that can be used for the structure determination of molecules with research interest. This type of radiation has a wavelength shorter than the visible light and much higher energy. The source of the X-rays has evolved over the years from vacuum tubes, where a metal anode would emit X-rays when bombarded with highly accelerated electrons, to synchrotron radiation and more recently to X-ray free electron lasers (XFEL).[1]

Synchrotrons and XFELs are based on the electromagnetic radiation emitted by charged particles (electrons) when acelerated.[1,13] In synchrotrons the electrons are kept in motion in a circular path (storage rings) and radiation is emitted, on a 90 degree angle, when their typical linear path is forced to change to circular.[13] On the other hand, with XFELs, the electrons are linearly accelerated in a kilometers-long system of superconducting cavities by an oscillating microwave field.[1] The accelerated electrons then pass through an undulator where they will emit electromagnetic radiation.[1]

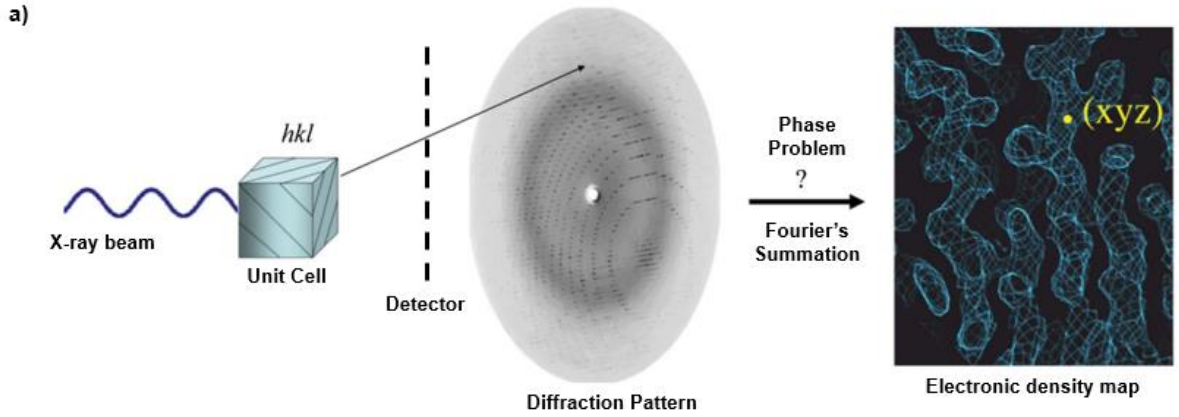
Besides X-ray diffraction, new methods have been developed to analyze macromolecules' crystals, such as neutron scattering. This technique allows the determination of the positions of hydrogens atoms in the structure, an essential part of enzyme mechanisms.[1,3] Information about the precise location of the hydrogen atoms is usually masked by the electronic cloud detected by the X-rays.[1]

Nowadays, X-ray crystallography is the only structural technique that allows the study of biological macromolecules at an atomic level, followed closely by Cryo-EM.[3] However, for optimal X-ray diffraction patterns, crystals have to be obtained from extremely pure and homogeneous protein solutions and also need to be of certain quality and size (at least 10μm[3] but preferably bigger) in order to diffract.[3,6] The smaller the crystal, the less diffracted beams (reflections) are observed. The use of synchrotron radiation has allowed the decrease of the size and number of crystals needed for diffraction, but the stability of the crystal still needs to be optimal.[3] Uniformity in the periodic arrangement of the crystal also correlates directly to the extent and quality of the diffraction pattern so the higher the degree of internal order of the crystal, the higher the atomic detail of the diffraction data.[1,3]

The unit cell, and its three-dimensional packing, isn't symmetrical from all sides. This means certain regions of the scattering pattern of the X-rays will be enhanced and others will be annulled (systematic absences) depending on the symmetry of the unit cell.[1] Therefore multiple

sides of the crystal must be exposed to the beam in order to get a full/complete set of reflections, representative of the crystal's structure.[1] The average number of measurements per unique reflection is called multiplicity or redundancy. [1] High multiplicity is ideal but means that the crystal will be exposed to the X-rays for longer periods of time or at higher intensities, resulting in radiation damage to the crystal and decreasing the quality of the diffraction pattern.[1] A good diffraction data set should have high resolution limit, high level of data completeness (*C*), high redundancy (*N*), low signal-to-noise ratio ($I/\sigma(I)$) and high space group symmetry.[2,14]

When a crystal is exposed to a monochromatic X-ray beam, every atom/electron of the three-dimensional crystalline lattice will cause the scattering of the x-ray beam.[1,15] Individual scattered rays can then be in phase or out of phase with the scattered rays of other atoms.[15] In crystalline lattices where the unit cell arrangement is well-defined, the different scattered rays will all be in phase with each other, obeying Bragg's law ($\lambda = 2d\sin\theta$).[1,15,16] If this is the case, the scattered rays are then classified as diffracted rays (or reflections).[15] After passing through the crystal, the diffracted rays reach a detector that will register their intensities.[16] After the phase problem is solved, the direct relation between the intensity and the amplitude of the wave ($|F_{hkl}|$) allows the determination of the electron density at a given position (Fourier's summation, Figure 3.b).[16]



b) $$\rho(xyz) = \frac{1}{V}\sum |F_{hkl}| \exp(i\alpha_{hkl}) \exp[-2\pi i(hx + ky + lz)]$$

Figure 3 - *a) The diffraction experiment. b) Equation of Fourier's summation. V is the volume of the unit cell and αhkl is the phase associated with the structure-factor amplitude |Fhkl|. Adapted from Figure 1 of reference 13.*

X-ray diffraction experiments result in a series of reflections, each characterized by amplitude and phase.[1] The amplitude is derived from the intensity of the diffraction scattering pattern, but the phase information gets lost in process.[1,16] However, both amplitude and phase

are needed to calculate the electron density maps that will lead to the protein's structure.[16] This is known as the phase problem. The solution to this problem is to indirectly estimate the phase by one of three types of methods: direct, molecular replacement (MR) or isomorphic substitution (derivatization).[1] In direct methods, typically used for small-molecule resolution, an atomic resolution of at least 1.2Å is needed and a relation between structure factors is made according to their probability.[1]

For macromolecules, however, that level of resolution is not always easy to achieve and the number of atoms to consider starts to be very high, so MR and heavy-atom variations are the most commonly used.

With MR, it's possible to derive the phase using the known atomic coordinates of structurally similar and homologous proteins (using Fourier transformation and Patterson's function).[1,16] This is only possible when the sequence identity between our target protein and the homologue used as a search model is higher than 20% and there's a root-mean-square deviation (RMSD) of the atomic positions of less than 2.0Å.[16] This method requires a lot of approximations by the resolution software so the chosen model needs to be manually refined in order to reach correct parameters, specific to our model.[1] Since the structure of one of the proteins used in this project had already been solved and published (PDB ID 6R62), MR was the method used for structure determination in this project.

When there's no homologous structure available, isomorphic substitution can be used, namely multiple isomorphous replacement (MIR), and single or multiple-wavelength anomalous diffraction (SAD and MAD, respectively). With MIR, the crystals are soaked in heavy-atom solutions creating isomorphous derivatives with the same unit cell and orientation of the "native" crystals. Both native and heavy-atom substituted crystals are exposed to X-rays and the intensity changes from one diffraction pattern to the other allow the deduction of the position of the heavy-atoms.[1,16] Single/multiple-wavelength anomalous diffraction (SAD/MAD) depends on the detection of the irregular scattering of heavy-atoms introduced into the protein's sequence by producing a recombinant variation of the desired protein with selenomethionine instead of "normal" methionine, for example.[16] Selenium has a characteristic scattering pattern, allowing the determination of its exact location in the structure.[1,16]

After data collection and data solution comes structure refinement, where the structural model (structure factors) is fitted to the experimental data (electron density maps).[14,17] This is an interactive process, done through several cycles of minimization of the discrepancies between

the calculated model and the experimental data, as well as model building of ligands, solvent molecules, ions, alternative side chain conformations and others, until all the electron density of the structure is justified.[15]

The visualization of the structural details of the model throughout the refinement process is done with electron density maps. The most used ones are maximum-likelihood σ_A weighted maps, which show a map with model bias correction ($2mF_o$-DF_c) and a difference map (mF_o-DF_c), showing incorrect densities in the model. In the map equations, *m* represents the figure of merit and D the Luzzati coefficient.[2]

One of the parameters used to determine the global quality of the model throughout the refinement is the R-factor. The crystallographic R-factor (or R-work), obtained with Equation 1, correlates the amplitudes of the structure factors calculated from a crystallographic model (F_c, the theoretical data) and those derived from the X-ray diffraction data (F_o, the obtained data).[17]

$$R = \frac{\sum \left| |F_o| - |F_c| \right|}{\sum |F_o|}$$

Equation 1

Since the goal of model refinement and optimization is to minimize the differences between the real data (F_o) and the theoretical model (F_c), the R-factor is not a fully independent parameter.[17] Minimization of F_o-F_c will lead to a decrease of the R-factor value, regardless of the refinement's accuracy.[17] For example, basing the refinement process only on the R-factor could easily lead to overfitting with the use of too many refinement parameters (e.g. adding too many water molecules with no corresponding electron density).

As a solution, a variation of the R-factor was introduced in 1992: the R-free, which corresponds to the R-factor of a set of randomly selected reflections that are left out of the refinement process, remaining "free" of any user/refinement error (typically 5% of the total number of reflections).[17] Thus, by comparing the difference between the R-work and the R-free values, we can determine the phase accuracy of the atomic model. Generally, if the difference between the two R-factors is higher than 5%, it shows a case of overfitting.[17]

Although important, the R-factors should not be used as the sole indicators of the correctness of the refinement.[17] An accurate refinement, and a valid model, should have no unjustified strong intensities peaks, make biological sense and be chemically reasonable regarding bond lengths, bond angles, atom coordination and geometry.[15]

1.2. Class II pyruvate aldolases

The basis of all organic compounds are the carbon-carbon bonds that form them. The industrial synthesis of organic compounds is extremely pollutant, hard and expensive but luckily, nature found a much more productive and simple way of synthesizing organic compounds: enzymes.[18] Enzymes are proteins with catalytic activity/properties, present in all living organisms, catalyzing a myriad of reactions and products.

Aldolases are enzymes capable of catalyzing carbon-carbon bond formation reactions, in a reversible and stereospecific manner, making them a great tool for the synthesis of organic molecules at the industrial level.[19] Aldolases usually have very high specificity regarding the acceptor substrate and are more flexible regarding the donor, for donor substrate with the same moiety.[18] However, aldolases show a diverse spectrum in terms of sequence, structure and catalytic residues involved in the reaction.[20]

These proteins can be divided in two main classes: class I aldolases, which require a lysine residue to form a Schiff base intermediate and generate a enamine nucleophile with the keto function of the substrate; and class II aldolases, which requires a divalent metal ion in the active site in order to promote the enolization of the donor substrate via Lewis acid complexation.[18,20,21] Regardless of the intermediate formed, the nucleophilic enamine or enolate attack the carbonyl carbon of the acceptor substrate, forming the new C–C bond.[18]

Inside each class, aldolases are further classified according to the specific nucleophilic substrate they use, with the main substrates being dihydroxyacetone phosphate (DHAP), pyruvate/2-oxobutyrate and acetaldehyde.[18,21–23] In this project, we'll be focusing on class II pyruvate aldolases.

Class II pyruvate aldolases are commonly found in bacteria, fungi, and algae, typically displaying an $(\alpha/\beta)_8$ barrel fold.[23–25] The active site is typically located near the C-terminal end of the beta strands that form the core of the barrel.[26] These enzymes are usually homo-hexamers in their active conformation, characterized by the existence of an inter-swapped domain between dimers (Figure 4).

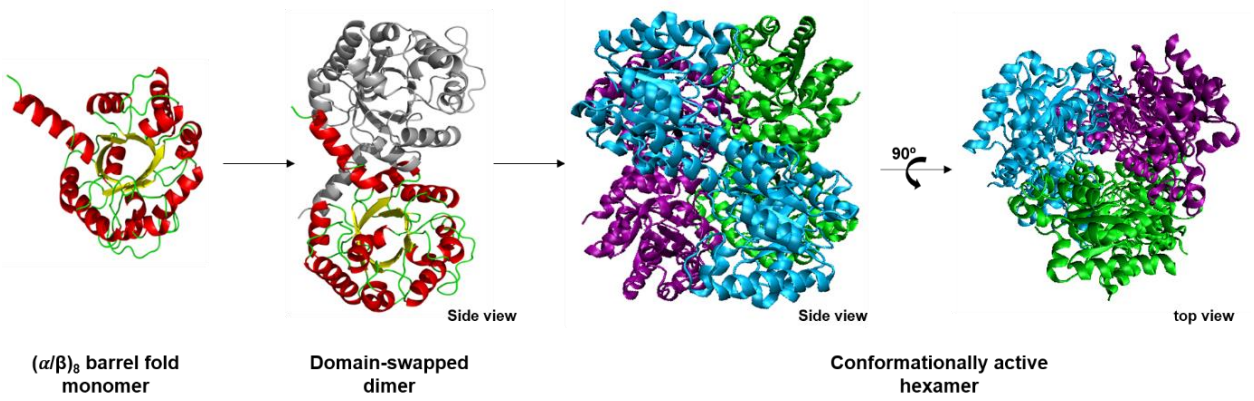


Figure 4 - *Representation of the general structure of class II pyruvate aldolases (PDB ID 6R62).*

However, the catalytic dimer does not correspond to the inter-swapped dimer as the active site is located in-between two other monomers in the hexamer (Figure 5). This hexamer corresponds to the minimum catalytic entity.

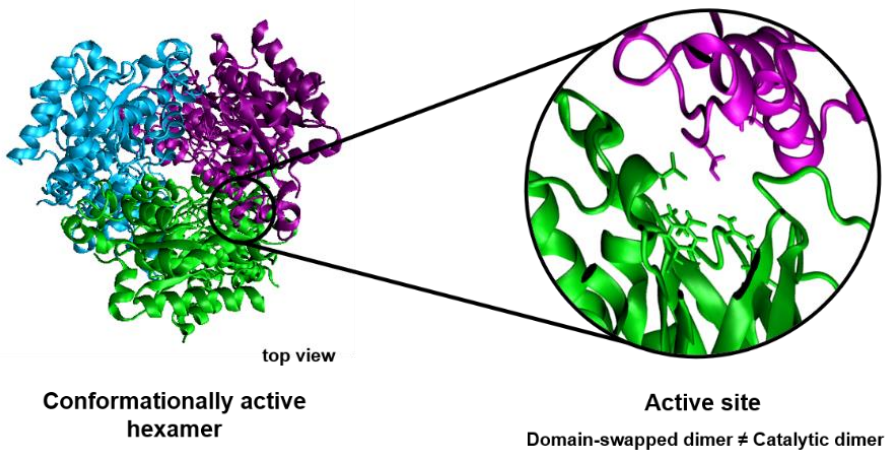


Figure 5 - *Representation of the active site location of class II pyruvate aldolases (PDB ID 6R62).*

In this project, two different class II pyruvate aldolases were studied: HpcH/HpaI aldolase from *Sphingomonas wittichii* RW1 (*Sw*HKA, UniProt ID: A5VH82) and 2-dehydro-3-deoxyglucarate aldolase from *Paraburkholderia phytofirmans* (*Bp*HKA, UniProt ID: B2T1L6). *Sw*HKA and *Bp*HKA are 251 and 260 residues long, respectively, and show a sequence identity of 31.5% between each other. Although the proteins show low homology at the amino acid level, they catalyze the same reaction, differing only in the configuration of the 4C stereocenter.

These proteins were identified as promising targets for the catalysis of aldol reactions by *De Berardinis et al.* in 2017, in a study that showed the enzyme's capability to use both pyruvate and hydroxypyruvate as the donor substrate.[27] This tolerance regarding the nucleophilic substrate

may seem irrelevant, since pyruvate and hydroxypyruvate only differ in one hydroxyl group, but the ability to catalyze more than one specific donor is rare in this type of proteins and would allow the synthesis of a much larger spectrum of ketoacid products.[28]

Focusing on hydroxypyruvate (HPA) as the nucleophilic donor substrate, *Sw*HKA catalyzes the reversible aldol coupling of HPA and an acceptor aldehyde, forming a 3(*S*),4(*S*)-dihydroxy-*α*-ketoacid (Figure 6).[29,30] *Bp*HKA catalyzes the same reaction, also reversibly, but the product is a 3(*S*),4(*R*)-dihydroxy-*α*-ketoacid (Figure 6).[30] Due to the specific products *Sw*HKA and *Bp*HKA synthesize, they can be further classified as class II hydroxy ketoacid aldolases (abbreviated to HKA).

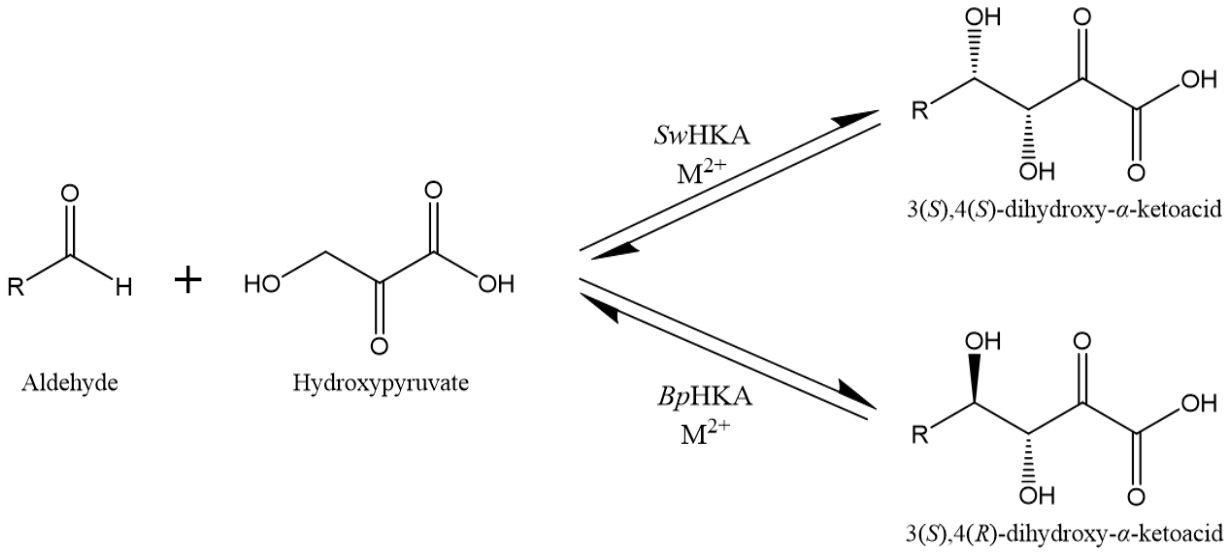


Figure 6 - *Reaction mechanism of SwHKA and BpHKA with hydroxypyruvate as the nucleophilic donor substrate. Adapted from reference 30.*

Biologically speaking, *α*-keto acids, where the keto group (C=O) is adjacent to the carboxylic acid group (COOH), are particularly important due to their involvement in processes like glycolysis and the Krebs cycle.[31] More specifically to this project, 3,4-dihydroxy-*α*-ketoacids are of great interest for the pharmaceutical and food industries but the synthesis is synthetically challenging.[28,29] The use of enzymes like *Sw*HKA and *Bp*HKA could be a great alternative for the industrial synthesis of these compounds.

2. Methods and materials

All experiments were done at the Sample and Preparation Characterization (SPC) lab at EMBL-Hamburg. The work reported in this thesis is in collaboration with the Biocatalysis group of Professor Ulf Hanefeld at the Delft University of Technology (TU Delft, Netherlands), and they were responsible for the expression and purification of all protein samples. The starting point with all protein samples was a purified protein sample in 20mM HEPES (pH 7.5).

2.1. General procedures

Before each of the following experiments was done, the protein solution was centrifuged at 9279 RCF x g, for 2 minutes (*Centrifuge 5415 R*, Eppendorf).

Protein concentrations were always determined with *Nanodrop 2000* (ThermoScientific), using the respective buffer mix as the blank.

When needed, final volumes were adjusted with type I ultrapure water (*PURELAB flex*, ELGA). All water used in this report came from this purifying system.

2.2. Quality control: SDS-PAGE

As a quality control measure, SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) gels were done for all protein samples, except *Sw*HKA-WT which was well studied by the start of this project.

In this technique, a detergent (SDS) is added to the protein solution and the mixture is heated up at high temperatures in order to denature the protein. Upon denaturation, the protein goes from the folded form to the linear one, exposing the polypeptide core. The SDS will bind to the polypeptides and break non-covalent interactions, evening out the differences in charge. This way, when an electrical field is applied to the gel, the protein's migration speed towards the anode only depends on its molecular weight.[32]

Before the gel was loaded, 10 µL solutions of each protein at 1 mg/mL with 2.5 µL of *NuPAGE*® LDS Sample Buffer (4X) (novex® by life technologies) and 1 µL of the reducing agent 2-mercaptoethanol (BME) at 1 mg/mL were heated at 70ºC for 10 minutes. After heating, the solutions were centrifuged and loaded into a 1.0mm x 15 well *NuPAGE*® 4-12% Bis-Tris Gel

(invitrigen by Thermo Fisher Scientific). Thermo Scientific's PageRuler Plus Prestained Protein Ladder was also loaded into one of the wells (5 µL). The gel was set to run for 45 minutes at 150 V (*PowerPac*™ Basic Power Supply, Bio-Rad) with 1X MES as the running buffer. After the run was over, the gel was stained overnight with *InstantBlue*™ (expedeon) and posteriorly washed several times with type I ultrapure water (*PURELAB flex*, ELGA).

2.3. Thermal Stability: nDSF assays

Thermal shift assays are useful methods of determining the stability of a protein when exposed to different conditions (e.g. pH values, buffers and additives).[5] In this report we used a differential scanning fluorimetry (nDSF) technique, which relies on the intrinsic fluorescence of a protein's tryptophan residues, instead of using a fluorescent dye. As the temperature increases, the protein starts to unfold, exposing the TRP residues in the hydrophobic core and allowing the detection of a fluorescent signal. Ideally, the transition between the folded and unfolded stages, after a certain temperature, results in a sharp increase in fluorescent signal over a short temperature range. This transition can be represented by a sharp sigmoidal curve (Figure 7) that can be used to calculate the protein's melting temperature (T_m) at 50% unfolded fraction.[5] Higher T_m values are associated with higher structural order and less conformational flexibility, correlating high Tm values with protein stability.

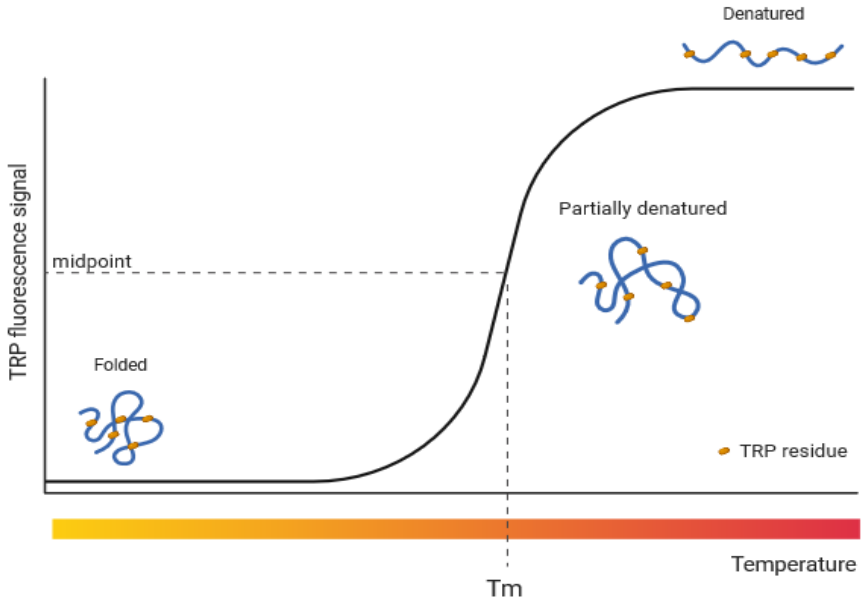


Figure 7 - *Typical differential scanning fluorimetry (DSF) assay by thermal denaturation. Figure adapted from a BioRender template.*

All experiments were done with Prometheus NT.48 (Nanotemper Technologies), between 15 and 95ºC, with a temperature increase rate of 1ºC/minute and at a fluorescence intensity of 50%. Prometheus NT.48 is an extremely precise dual-wavelength system that measures fluorescence emission at both 330 and 350nm, the wavelengths that describe the tryptophan emission shift upon unfolding.[33,34] It is hard to determine T_m values from single wavelength measurements but an apparent T_m value can be derived from the inflection point of the sigmoidal graph of the 350/330nm ratio or the scattering, determining a well-defined transition between the folded and unfolded stages of the protein.[33,34]

The characterization and overall stability of both *Sw*HKA (WT and mutants) and *Bp*HKA was already known in the lab, so the focus was on the stability of the proteins in the presence of the magnesium cofactor, hydroxypyruvate and different aldehydes.

Based on previous knowledge of this system and on information reported in the literature for homologues proteins, 70 nDSF assays were performed with *Sw*HKA and 51 with *Bp*HKA, using different conditions with: 0 to 1.9M $MgCl_2$; 0 to 275mM hydroxypyruvate (HPA); 20 and 50mM monosodium phosphate; 20 and 50mM monopotassium phosphate; 100mM DL-glyceraldehyde; 100mM glycolaldehyde and 100mM butyraldehyde. All proteins were used at a final concentration of 0.5mg/mL. The general incubation time of the samples was 45 minutes before the measurement, but incubation times of 2, 24 and 48h were also tested.

The stock solutions of these reagents are listed in the Supplementary data (S1). The best results were obtained with the conditions described on Table 1 for *Sw*HKA and Table 2 for *Bp*HKA.

Table 1 - *Custom nDSF assays of SwHKA. Assays done in triplicate. The magnesium ion (Mg) was added in the form of $MgCl_2$. The samples were incubated for 45 minutes before measurement.*

<table>
<tr><th>SwHKA</th><th colspan="2">Condition</th></tr>
<tr><td>WT</td><td colspan="2" rowspan="3">Blank (sample solution: 20mM HEPES, pH=7.5)</td></tr>
<tr><td>S116A</td></tr>
<tr><td>S116C</td></tr>
<tr><td>WT</td><td colspan="2" rowspan="3">20mM Mg</td></tr>
<tr><td>S116A</td></tr>
<tr><td>S116C</td></tr>
<tr><td>WT</td><td colspan="2" rowspan="3">20mM Mg + 50mM HPA</td></tr>
<tr><td>S116A</td></tr>
<tr><td>S116C</td></tr>
<tr><td rowspan="7">WT</td><td colspan="2">stored at 4ºC for 2 weeks</td></tr>
<tr><td colspan="2">stored at -80ºC</td></tr>
<tr><td>0, 10, 25, 50, 75, 100, 150, 200, 250, 300, 400 and 500mM</td><td>Mg</td></tr>
<tr><td>0, 10, 25, 50, 75, 100, 150, 200, 250 and 275mM</td><td>HPA (+200mM Mg)</td></tr>
<tr><td colspan="2">100mM Mg + 100mM HPA + 100mM DL-Glyceraldehyde</td></tr>
<tr><td colspan="2">100mM Mg + 100mM HPA + 100mM Glycolaldehyde</td></tr>
<tr><td colspan="2">100mM Mg + 100mM HPA + 100mM Butyraldehyde</td></tr>
</table>

Table 2 - *Custom nDSF assays of BpHKA. Assays done in triplicate. The magnesium (Mg) and sodium (Na) ions added in the form of $MgCl_2$ and NaCl, respectively. The samples were incubated for 45 minutes before measurement, unless stated otherwise.*

***Bp*HKA**	**Condition**	
WT	Blank (20mM HEPES, pH=7.5)	
	20mM Mg	
	20mM Mg + 50mM HPA	
	50mM Mg	
	50mM Mg (24h)	
	50mM Mg (48h)	
	100mM Mg	
	100mM Mg (24h)	
	100mM Mg (48h)	
	50mM Mg + 50mM Na	
	100mM Mg + 50mM Na	
	100mM Na	
	10, 25, 50, 75, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800 and 900mM 1, 1.1, 1.5, 1.75 and 1.9M	Mg
	0, 10, 25, 50, 75, 100, 150, 200, 250 and 275mM	HPA (+200mM Mg)
	100mM Mg + 100mM HPA + 100mM DL-Glyceraldehyde	
	100mM Mg + 100mM HPA + 100mM Glycolaldehyde	
	100mM Mg + 100mM HPA + 100mM Butyraldehyde	

2.4. Crystallizations screens and trials

At the beginning of this project, the crystallization conditions of *Sw*HKA-WT were already known: 0.5-0.8M sodium citrate and 0.1M HEPES at pH 7.0, at 277.15K (4ºC).[29] These conditions also proved successful in the crystallization of other mutants of *Sw*HKA-WT, consistently resulting in bipyramidal crystals whose diffraction improved with the aging of the crystals.

However, when approaching the crystallization of the S116A and S116C mutants, finding different crystallization conditions was desirable since the citrate, found in previously identified crystallization conditions, tended to complex with the magnesium in the active site, preventing the binding of the hydroxypyruvate. Conditions at room temperature would also be preferable.

Regarding *Bp*HKA, some crystallization hits were already obtained but these were not reliable or reproducible, so the search for better crystallization conditions was also a goal.

With the aim of identifying a set of completely different crystallization conditions, commercially available crystallization screens were tested to guide the determination of the crystallization conditions of the S116A and S116C *Sw*HKA mutants and of the *Bp*HKA-WT (Table 3).

Table 3 - *Commercially available screens used. In the BpHKA screens the base reservoir solution was composed of 8% PEG2000-MME and 0.1M $MgCl_2$.*

Protein	[P] (mg/mL)	Screen	Supplier	Total nº of conditions tested
*Sw*HKA-S116A	7	Classic I Suit	Qiagen	192
*Sw*HKA-S116C				
*Bp*HKA-WT	5	Additive Screen	Hampton Research	288
	7			
	8			

The screens were set up with the crystallization robot *Mosquito® LCP* (ttplabtech), on 96-well 3-drop *Swissci* plates (ttplabtech), using 100 nL of protein solution mixed with 100 nL of reservoir solution (sitting-drop technique). The plates were kept at 19ºC and loaded to *The RockImager-1000* (Formulatrix) where every drop in the plate was inspected at a standard frequency (0, 1, 3, 5, 7, etc days). The images of each drop were then uploaded to SPC's user website (CRIMS©, https://htx.embl-hamburg.de/), allowing the monitorization of the crystallization process. To guarantee that the "hits" were protein crystals, UV-Vis inspections were also done. If

the crystals observed were in fact from protein, they should glow under UV light inspection due to the fluorescence generated by protein's aromatic residues.

The hits were replicated and further optimized by hand, using *EasyXtal 15-Well Tool X-Seal* plates (Qiagen) and the hanging-drop technique. Drops of 1.5 µL of protein solution were mixed with 1.5 µL of reservoir solution and equilibrated against 0.5 mL reservoir solution. The reservoir solutions were either set up by hand or using the *Scorpion Screen Builder* (Art Robbins Instruments). The crystallization conditions used for *Sw*HKA-S116A and *Sw*HKA-S116C are described in Table 4.

Table 4 - *Crystallization conditions tried for SwHKA-S116A and SwHKA-S116C. Total of 108 manually set drops.*

***Sw*HKA**	**[Protein] (mg/mL)**	**Condition**	**P:RS ratio (µL)**	**Temperature (ºC)**	**Total nº drops**
S116A S116C	7	0,5-0.9M sodium citrate 0,1M HEPES (pH 7.0)	1.5 : 1.5	4	16
	7	25-35% PEG400 0,1M HEPES (pH 7.0-8.0) 0.1M and 0.2M $MgCl_2$	1.5 : 1.5	19	60
	7	20% and 23% PEG400 0,1M HEPES (pH 7.5 and pH 8.0) 0 - 0.2M $MgCl_2$	1.5 : 1.5	19	12
S116A	7	20% and 23% PEG400 0,1M HEPES (pH 8.0) 0.1M and 0.2M $MgCl_2$	1.5 : 1.5	19	8
S116C	5				
S116A	7	20% PEG400 0,1M HEPES (pH 8.0) 0.1M $MgCl_2$	1.5 : 1.5	19	6
			1.2 : 1.8		
			1.8 : 1.2		
S116A	7	18% PEG400 0,1M HEPES (pH 8.0) 0.1M $MgCl_2$	1.5 : 1.5	19	6
			1.2 : 1.8		

For *Bp*HKA, besides the hits from the commercially available screens, some conditions were curated based on the published crystallization conditions of a HpcH/HpaI aldolase/citrate lyase family protein from *Burkholderia cenocepacia J2315*, (PDB ID: 4MF4), which has 75% sequence identity to *Bp*HKA. 4MF4's crystallization conditions are: 20% PEG3350, 0.2M AmCl, 289K (15ºC); vapor diffusion method and sitting-drop technique. The use of "chinese" hanging-drop also proved to have positive effects in the crystallization process of *Bp*HKA, with some crystals only appearing in this type of drop. The "chinese" drop is a drop in which a protein solution is slowly released from the pipette creating a line on the cover slip. The reservoir solution/mother liquid is then added at the end of the protein line, which results in a concentration gradient, adding dynamics to the system. From a total of 570, the most promising crystallization conditions used for *Bp*HKA-WT are described in Table 5.

Table 5 - *Crystallization conditions of BpHKA. Total of 326 drops. All drops were manually set except the conditions with a protein(P):reservoir solution(RS) ratio of 1:1, which were set with Mosquito®LCP. Drops marked with an asterisk (*) represent conditions where both "chinese" and "normal" drop set-ups were used.*

[Protein] (mg/mL)	**Condition**	**P:RS ratio (µL)**	**Temperature (ºC)**	**Total nº drops**
20.8	20% PEG1500 0,1M bis-tris propane (pH 8.5) 0,1M $MgCl_2$	1.2 : 1.8	19	6
		1.5 : 1.5		
10.4		1.8 : 1.2		
20.8	20-24% PEG1500 0,1M bis-tris propane (pH 7-8.5) 0-0,1M $MgCl_2$	1 : 1	19	60
	20-24% PEG1500 0,1M $NaHPO_4$ buffer (pH 6-7.1) 0-0,1M $MgCl_2$			36
20.8	12-18% PEG1500 0,1M bis-tris propane (pH 8.5) 0,1M $MgCl_2$	1.8 : 1.2	19	6
10.4				
10	16-17% PEG1500 0,1M bis-tris propane (pH 7-8.5) 0-0,1M $MgCl_2$	1.8 : 1.2	19	2

[Protein] (mg/mL)	Condition	P:RS ratio (µL)	Temperature (ºC)	Total nº drops
9	18-24% PEG1500 0,1M bis-tris propane (pH 7-8.5) 0-0,1M $MgCl_2$	1 : 1	19	48 48
11	20-24% PEG1500 0,1M bis-tris propane (pH 7-8.5) 0-0.1M MgCl2	1 : 1	19	48
6 8	14-16% PEG1500 0,1M bis-tris propane (pH 8.5) 0,1M $MgCl_2$	1.5 : 1.5	4	4
10	5-10% PEG400 16% PEG1500 0,1M bis-tris propane (pH 8.5) 0,1M $MgCl_2$	1.5 : 1.5	19	2
10	16% PEG1500 0,1M bis-tris propane (pH 8.5) 0.1-0,25M $MgCl_2$	1.5 : 1.5	19	4
8	20% PEG3350 0,2M $MgCl_2$	1.5 : 1.5	19	1
8 4	10-18% PEG3350 0,2M $MgCl_2$	1.5 : 1.5	19	10
8	4-8% PEG3350 0,2M $MgCl_2$	1.5 : 1.5	19	3
4	12% PEG3350 0.05-0.15M $MgCl_2$	1.5 : 1.5	19	3
7	10-18% PEG1500 0,1M HEPES (pH 8.5) 0,1M $MgCl_2$	1.5 : 1.5	19	5

[Protein] (mg/mL)	Condition	P:RS ratio (µL)	Temperature (ºC)	Total nº drops
10	6-10% PEG3350 0,2M $MgCl_2$	1.5 : 1.5	19	3
8 10	8.5-9% PEG3350 0,2M $MgCl_2$	1.5 : 1.5	19	4
9.5	25% PEG400 25% PEG550-MME 8% PEG2000-MME 8% PEG4000 8% PEG5000-MME 0-0,2M $MgCl_2$	1.5 : 1.5	19	15
9.5	8% PEG2000-MME 0,2M $MgCl_2$	1.5 : 1.5	19	11*
7	8% PEG2000-MME 0,1M $MgCl_2$	1.5 : 1.5	19	2*
7 8	8% PEG2000-MME 0,1M HEPES (pH 7.5-8) 0,1M $MgCl_2$	1.5 : 1.5	4	8*
7	8% PEG2000-MME 0,1M $MgCl_2$ 4% formamide	1.5 : 1.5 1.2 : 1.8	19	4*
7	8% PEG2000-MME 0,1M $MgCl_2$ 4% PEG400	1.5 : 1.5 1.2 : 1.8	19	3*
9.5	10-18% PEG5000-MME 0,1-0.2M $MgCl_2$	1.5 : 1.5	19	10
7	14-22% PEG5000-MME 0,2M $MgCl_2$	1.5 : 1.5	19	5

For all proteins in this report, when appropriate-sized crystals were grown, these were soaked in relevant soaking solutions (41 in total), consisting of different combinations of: 20 to 200mM $MgCl_2$; 0 to 150 and 200mM HPA and 0 to 500mM DL-glyceraldehyde. Soaking times mainly varied between 0 (flash soak), 5, 10, 15, 20, 30, 40, 45, 50, 60 and 90 minutes, but overnight soaks up to 24h were also tried when dealing with both substrates (HPA and DL-glyceraldehyde) at the same time. Besides the cofactor, the substrates, and the respective reservoir solution of each crystal, a cryoprotectant solution was also always added to the soaking solutions. Two cryoprotectants were tested, 20% glycerol and 30% PEG400, and PEG400 proved to be the best cryoprotectant for all crystals.

As stated before, *Sw*HKA-WT crystals grew in a citrate solution, which competes for the magnesium in the active site, forming magnesium citrate. To prevent this from happening, before soaking the crystals in the relevant soaking solution (containing magnesium), an established protocol was done, where the citrate-containing mother liquid (sodium citrate and HEPES at pH 7.0) was slowly exchanged over a period of 30 minutes, at 4ºC, to a higher concentration potassium bromide mother liquid (potassium bromide, HEPES at pH 7.0 and the cryoprotectant PEG400). After the exchange, the soaking of the *Sw*HKA-WT crystals was done as described in the previous paragraph.

After the soaking procedure, the crystals were manually harvested using either nylon or lithographic cryo-loops (Hampton Research or MiTeGen) of the appropriate size. After manual harvesting, the crystals were immediately flash frozen in liquid nitrogen and stored in SPINE pucks (*SC3 Baskets*, MiTeGen), inside storage dewars at cryo temperatures until x-ray diffraction data collection was possible.

2.5. Final crystallization and soaking condition

2.5.1. *Sw*HKA

Protein crystals of S116A-Mg (*holo*-S116A) were obtained at 292K using the vapor diffusion method, from drops resulting from mixing 1,8µl of a protein sample at 7 mg/ml (in 20mM HEPES, pH 7.5) with 1,2µl of crystallization solution consisting of 30% PEG400, 0.1M $MgCl_2$ and 0.1M HEPES at pH 7.50. Thin triangular crystals took 2-5 days to appear and remained consistent in size and shape as they grew older.

Crystals of S116A-Mg-HPA were obtained by transferring crystals of S116A-Mg into a new drop containing 30% of PEG400, 0.1M $MgCl_2$, 0.1M HEPES at pH 7.50 and 50 mM of hydroxypyruvate (HPA). The crystals were left in this new condition for 15 minutes and then were flash frozen in liquid nitrogen.

*Apo-Sw*HKA crystals were grown according to the established protocol: vapor diffusion method, at 277.15K, drops with 1,5µl of a protein sample at 7 mg/ml (in 20mM HEPES, pH 7.5) and 1,5µl of crystallization solution consisting of 0.5M sodium citrate and 0.1M HEPES at pH 7.0. Bipyramidal crystals took 3-5 days to appear. After 21 days of aging, a citrate/KBr mother liquid exchange was performed and the crystals were flash frozen in liquid nitrogen. The crystals remained consistent in size and shape as they grew older.

Although the structures of *Sw*HKA-Mg (*holo-Sw*HKA) and *Sw*HKA-Mg-HPA will be used in this report for the analysis of the other *Sw*HKA and *Sw*HKA-S116A structures, these two structures had already been determined and studied by Dr. Isabel Bento, so there was no need to determine them again. The structure of *Sw*HKA-Mg has been submitted to the PDB but is being kept on hold waiting for publication. The structure of *Sw*HKA-Mg-HPA is already public, with PDB ID 6R62.

Crystals of *Sw*HKA-Mg-HPA-3GR, from now on referred to as *Sw*HKA-all, were obtained by transferring crystals of *Apo-Sw*HKA, grown in 0.55M sodium citrate and 0.1M HEPES at pH 7.0, into a new drop containing 0,6M KBr, 0.1M HEPES at pH 7.0, 30% PEG400, 0.1M $MgCl_2$, 0.1M hydroxypyruvate (HPA) and 0.210M DL-glyceraldehyde (equivalent to 0.105M D-glyceraldehyde (3GR)). The crystals were left in this new condition for 45 minutes and then were flash frozen in liquid nitrogen.

2.5.2. *Bp*HKA

Crystals of *Bp*HKA-Mg were obtained at 292K using the vapor diffusion method, from drops resulting from mixing 1,5µl of a protein sample at 8 mg/ml (in 20mM HEPES, pH 7.5) with 1,5µl of crystallization solution consisting of 8% PEG2000MME and 0.1M $MgCl_2$. A small rod-like crystal appeared after 9 days. A crystal was transferred into a new drop containing 8% PEG2000MME, 30% of PEG400 and 0.1M $MgCl_2$. The crystals were left in this new condition for 20 minutes and then were flash frozen in liquid nitrogen.

2.6. X-ray data collection

X-ray diffraction data was collected at beamlines P13 and P14 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). At the time, P13 was equipped with a DECTRIS PILATUS 6M detector (25Hz) and a MD2 diffractometer (ARINAX), resulting in data collection at 0.9763Å.[35] P14 was equipped with a DECTRIS EIGER 16M detector (133Hz) and a MD3 diffractometer (ARINAX), resulting in data collection at 0.9766Å.

Data collection was done at 100K, usually for a total of 3600 x-ray diffraction data images, using an oscillation range of 0.1º or, less often, for a total of 2400 images when using an oscillation range of 0.15º. This difference depended of the space group of the crystal (when previously known). Overall, 176 crystals were exposed to the x-ray beam, resulting in varying degrees of success: no diffraction, diffraction with ice rings and high resolution “clean” diffraction.

2.7. X-ray data processing

Data processing began with data reduction, using the XDS (X-ray Detector Software), a program for processing single-crystal monochromatic diffraction data.[36]

This was followed by data scaling, with Aimless (part of the CCP4 Software Suite bundle), which put all the collected data on a common scale.[37,38]

2.8. Structure solution

After scaling, the MOLREP software was used for the phasing.[39] All structures from *Sw*HKA (WT and S116A) and *Bp*HKA were solved by molecular replacement method (MR), using a previously obtained single-wavelength anomalous diffraction (SAD) *Sw*HKA model.

The refinement process consisted of several cycles of refinement (translation-libration-screw-rotation (TLS) and non-TLS) using REFMAC (5.8.0267 version), a program of the CCP4 suite for maximum-likelihood-based macromolecular refinement, and model building with *Coot*.[40,41] The use of the CCP4 Software Suite bundle was done through the CCP4 Cloud and CCP4i2.[38,42] When the structure model was sufficiently refined, ligands, solvent molecules and

alternate amino acid configurations were added to the system where there was justifiable electron density and the refinement process continued until convergence of the R and R_{free} factors.

2.9. Other techniques

Other techniques were learned during the laboratory phase of this master thesis, but as the obtained results were not relevant for this report they were not included. Nevertheless, a list of those techniques is mentioned below.

Besides manual crystal harvesting/phishing, automated harvesting was also done using the automatic crystal harvester CrystalDirect® (ARINAX). This technique was used with *Bp*HKA and the set-up procedure was done with the crystallization robot *Mosquito® LCP* (ttplabtech), as described before, using 96-well CrystalDirect™ plates instead of the 96-well 3-drop *Swissci* plates. However, these crystallization trials were unsuccessful, either due to the crystallization conditions or the different equilibrium kinetics caused by the low drop volume (200nL).

In this report the main crystallization method used was vapor diffusion, both by the hanging-drop (when done manually) and sitting-drop techniques (when done with the *Mosquito® LCP*), but some batch method, co-crystallization and seeding experiments were also used/done, mainly with *Sw*HKA, but with little success.

Mass photometry was also done with *Bp*HKA in order to confirm the hexameric conformation of the protein. This technique had already been made for *Sw*HKA, with similar results.

3. Results and discussion

3.1. Quality Control: SDS-PAGE

3.1.1. *Sw*HKA

As stated before, *Sw*HKA-WT was already being studied in the lab and an SDS-PAGE gel had previously been done with the protein sample, which proved to be pure and showed no signs of degradation. Therefore, an SDS-PAGE gel was done to analyze the degree of purity of both mutant samples. Figure 8 represents the SDS-PAGE gel obtained with the *Sw*HKA mutants S116A and S116C.

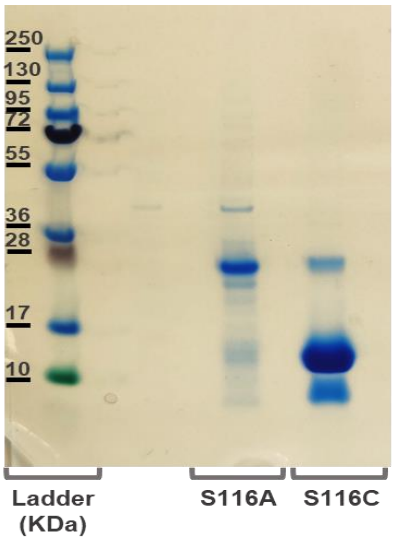


Figure 8 - *SDS-PAGE gel of S116A and S116C mutants of SwHKA. From left to right: protein ladder, S116A protein sample, S116C protein sample.*

In the S116A lane of Figure 8, there's a strong band at around 26KDa, which corresponds to the protein's molecular weight, suggesting that the S116A protein sample is quite pure and without much degradation. However, the same cannot be said for the S116C protein sample. In the S116C's sample lane (molecular weight also 26.7KDa), there are several bands indicating a high degree of degradation, with the most intense bands corresponding to molecular weight lower than 17KDa.

There are several possible explanations for the difference in the degradation profile of the protein samples of the two mutants. Although S116A and S116C only differ in one amino acid residue, this mutation can be enough to negatively affect the stability of the S116C mutant, resulting in a much less stable protein. However, there could've also been an error either in

S116C's production process or storage that caused the protein to degrade much faster than S116A and *Sw*HKA-WT.

3.1.2. *Bp*HKA

Regarding *Bp*HKA, an SDS-PAGE gel was also performed at the beginning of this project and the resulting band was similar to the one shown in Figure 9, on the right (labelled "0h"). The protein sample was deemed pure and the crystallization experiments began.

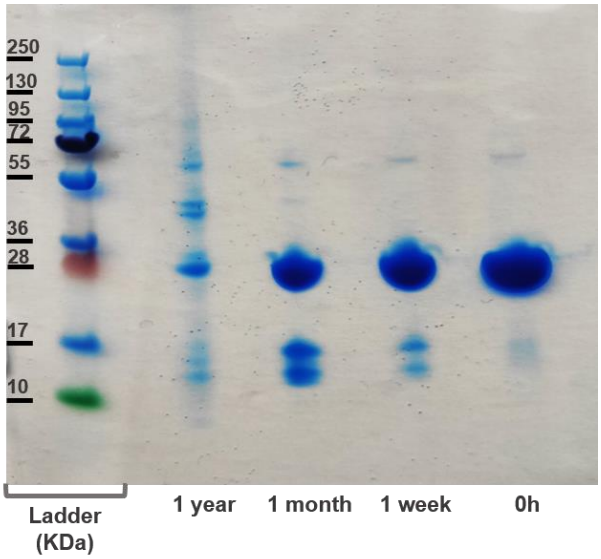


Figure 9 - *SDS-PAGE gel of BpHKA. The samples used were stored for different periods of time at 4ºC and at -80ºC. Time stored at 4ºC. From left to right: approximately 1 year, 1 month, 1 week and 0h (sample defrosted directly before the gel was set-up).*

However, as we'll see going forward, the crystallization process of *Bp*HKA was not always reproducible, even when replicating the exact same crystallization set-up and conditions that yielded promising crystals. After ruling out other possible causes for this unpredictability of crystallization results, such as quality of the stock solutions, temperature stability in the crystal-growth room and sequencing of the protein solution to guarantee there hadn't been any mishandling of the protein sample, a second *Bp*HKA gel was set up with protein that had been stored at 4ºC for different periods of time (Figure 9). In this new gel, a sample of *Bp*HKA stored at -80ºC, with no time at 4ºC, was also loaded, replicating the condition that had yielded the gel with no significant degradation at the beginning of this project.

Although the periods of time tested at 4ºC were a bit extreme (Figure 9), and the protein was only stored at 4ºC for 1-3 days at most, we can clearly see that even after 1 week at 4ºC

*Bp*HKA already shows some degree of degradation. Even though the most intense band corresponds to the protein's molecular weight (27.9KDa), we also see two bands at molecular weight lower than 17KDa and a less intense band at around 60KDa. These bands could be the result of hydrolysis of some of the protein's bonds and aggregation, in the case of the bands at lower and higher molecular weight, respectively.

There was no indication that the protein samples were unstable in the storage buffer (20mM HEPES at pH 7.5). Furthermore, there was never reproducibility problems in the crystallization process of *Sw*HKA that would have suggested an issue with protein sample stability, so we were not expecting it with *Bp*HKA and, at the beginning, it was not considered the cause of the crystal reproducibility problem. After this discovery, and to prevent the use of degraded protein sample, the protein that was stored in batches of 100µL at -80ºC was aliquoted in volumes of 10µL and 15µL, frozen and stored again at -80ºC. From this point on, all experiments were done with protein defrosted directly before the experiment began.

Since a metal cofactor is essential for the activity and stability of class II pyruvate aldolases, we decided to access the effect of the metal in the protein sample stability. A long-term incubation experiment was set-up with *Bp*HKA, where two protein samples were left at 4ºC, incubated with and without $MgCl_2$. Aliquots of each of the samples were then taken at different intervals of time and frozen at -80ºC. After 6 weeks of total incubation time of the last aliquots, the previous samples were defrosted and a gel was run in order to compare the degradation level of BpHKA with and without the presence of the cofactor of interest (Mg^{2+}). The resulting gel is shown in Figure 10.

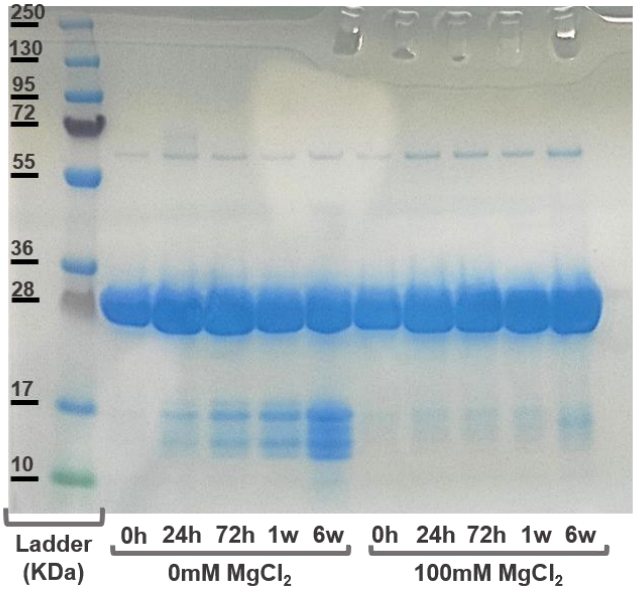


Figure 10 - *SDS-PAGE gel of BpHKA incubated with and without $MgCl_2$.*

Figure 10 represents an SDS-PAGE done with the protein sample incubated with 0mM and with 100mM $MgCl_2$. The samples incubated with magnesium do not show the same level of degradation overtime as the samples that weren't incubated with magnesium. The main band of every sample corresponds to *Bp*HKA's molecular weight (28.3KDa) and they all have a faint band at around 60KDa, which might correspond to aggregated protein. The biggest difference is at lower molecular weights, where samples without magnesium show two bands between 17 and 10KDa that start to show-up after 24h at 4ºC. These bands became stronger overtime, indicating that the protein's degradation increase with time. These bands seem non-existent in the samples incubated with 100mM $MgCl_2$, even after 6 weeks at 4ºC, which leads us to conclude that the magnesium ion does have a role in the stabilization of *Bp*HKA, preventing the hydrolysis of the protein bonds that would lead to bands at lower molecular weights.

Although minimal, the degradation of *Bp*HKA at 4ºC between 24 to 72h affected protein concentration, which in turn influenced the crystallization process of the protein and may be the reason underlying the reproducibility issues we were having. When the sample concentration was measured before the crystallization experiments, the concentration used didn't correspond to the active protein in solution. This may have affected the crystallization trials since the concentration used was not actually "real" and it varied between experiments, depending if the protein was freshly defrosted or had been stored at 4ºC.

3.2. Thermal stability: nDSF assays

3.2.1. *Sw*HKA

The melting temperature (T_m), temperature at which the concentration of the protein in the folded and unfolded states is the same, is a parameter of thermal stability that can be used to infer, to a certain degree, a protein's overall stability.[43] With the knowledge that the *Sw*HKA mutants showed lower enzymatic activity than *Sw*HKA-WT (manuscript in preparation), nDSF assays were performed in order to compare the thermal stability of *Sw*HKA-WT and the mutants S116A and S116C (Table 6).

Table 6 - *Melting temperatures (T_m) obtained through nDSF for the S116A and S116C mutants of SwHKA, as well as SwHKA-WT.*

Protein	Condition	T_m
WT	Blank (sample solution: 20mM HEPES, pH=7.5)	66.2°C
S116A		66.7°C
S116C		65.6°C
WT	20mM Mg	79.6°C
S116A		80.5°C
S116C		78.4°C
WT	20mM Mg + 50mM HPA	74.3°C
S116A		75.6°C
S116C		74.7°C

In Table 6, we see that in the *apo* state, the wild-type (WT) and mutant proteins have a similar T_m, leading us to believe they are similarly stable in this state. With the addition of magnesium to the system, the T_m values of the three proteins increased an average of 13ºC, which reflects a significant shift in protein stability. This increase in protein stability in the presence of magnesium is consistent with the properties of class II aldolases, whose activity depends on the presence of a metal cofactor.

When both magnesium and hydroxypyruvate (HPA) are added to the proteins, the T_m value decreases when compared to the temperatures with only magnesium in solution but are still considerably higher than the T_m's of the non-bound proteins (*apo* state), with an average difference of 8.7ºC.

Point mutations can impact the thermal stability of a protein, since amino acid interactions are temperature dependent, and vice-versa.[44] However, this does not seem to be the case of

*Sw*HKA mutants S116A and S116C, as the difference in T_m values between the WT and the mutants were negligible in the tested conditions. When looking at the three-dimensional structure of *Sw*HKA-WT (PDB ID 6R62), the mutated residue, serine 116, belongs in a small loop region that does not contribute to the overall fold of the protein. Therefore, mutating S116 to an alanine (S116A) or a cysteine (S116C) does not disrupt any secondary structure motifs, which consequentially does not significantly affect protein fold, protein stability and the T_m value.

In order to better understand *Sw*HKA and determine initial crystal soaking concentrations, nDSF assays were performed with *Sw*HKA-WT using a more extensive concentration range of $MgCl_2$ and HPA. This method was also used to test *Sw*HKA thermal behavior in the presence of 3 potential aldehyde substrates: DL-glyceraldehyde, glycolaldehyde and butyraldehyde. The resulting T_m values of these studies are shown in Table 7.

Table 7 - *Melting temperatures (T_m) obtained through nDSF for SwHKA-WT when incubated with 0 to 500mM of $MgCl_2$, 0 to 275mM of HPA and 3 different aldehydes.*

Condition		T_m
0mM	Mg	66.7°C
10mM		80.2°C
25mM		79.5°C
50mM		79.3°C
75mM		79.6°C
100mM		79.5°C
150mM		76.9°C
200mM		75.4°C
250mM		75.8°C
300mM		76.0°C
400mM		76.7°C
500mM		77.3°C
0mM	HPA (+200mM Mg)	75.6°C
10mM		75.9°C
25mM		76.2°C
50mM		76.4°C
75mM		76.6°C
100mM		76.7°C
150mM		77.1°C
200mM		76.1°C
250mM		78.0°C
275mM		76.1°C

Condition	T_m
100mM Mg + 100mM HPA + 100mM DL-glyceraldehyde	71.7°C
100mM Mg + 100mM HPA + 100mM glycolaldehyde	71.5°C
100mM Mg + 100mM HPA + 100mM butyraldehyde	71.9°C

If we focus on the T_m values when the protein was incubated with 0 to 500mM $MgCl_2$, we see that from 0mM to 10mM the T_m shifts from 66.7ºC to 80.2ºC, an increase of 13.5ºC, that reflects the immediate improvement of *Sw*HKA's thermal stability in the presence of the magnesium ion. The T_m values then seem to reach a peak from 10 to 100mM $MgCl_2$, with the highest registered T_m value at 10mM. From 150 to 500mM the T_m values decrease a little, about 3-4ºC, but remain much higher than the initial T_m of 66.7ºC at 0mM $MgCl_2$. These results appear to indicate that at magnesium concentrations higher than 150mM, SwHKA's stability starts to decrease. This could be due to oversaturation with the magnesium ion, which at concentrations higher than 100mM starts to disrupt/destabilize the protein, resulting in a decrease of the T_m.

In the presence of $MgCl_2$ and an increasing concentration of hydroxypyruvate (HPA), the shift in T_m values is smaller, increasing only about 2.4ºC between 0 and 250mM HPA. The T_m values obtained in these conditions are also similar to the values obtained with 200 to 500mM $MgCl_2$, indicating that the presence of HPA offsets the presence of higher concentrations of magnesium regarding *Sw*HKA's thermal stability, but does not generally improve it.

When comparing the T_m values resulting from incubation with the different aldehydes, all possible substrates of *Sw*HKA, we see that these values do not vary significantly, suggesting that, at least regarding its thermal stability, the protein does not show a preference for any of the aldehydes tested.

From the nDSF results of *Sw*HKA, the concentration of $MgCl_2$, HPA and DL-glyceraldehyde chosen for the soaking experiments was initially 100mM of each, although many more variations were used.

3.2.2. *Bp*HKA

Since *Bp*HKA was much less studied at the beginning of this project, the initial stability nDSF screening of this protein encompassed more conditions and incubation times than the ones used for *Sw*HKA (Table 8).

Table 8 - *Melting temperatures (T_m) obtained through nDSF for BpHKA-WT. Magnesium (Mg) and sodium (Na) ions were added in the form of $MgCl_2$ and NaCl, respectively.*

Condition	T_m
Blank (20mM HEPES, pH=7.5)	55.8°C
20mM Mg	61.8°C
20mM Mg + 50mM HPA	72.2°C
50mM Mg	64.1°C
50mM Mg (24h)	63.8°C
50mM Mg (48h)	63.9°C
100mM Mg	68.9°C
100mM Mg (24h)	68.0°C
100mM Mg (48h)	68.3°C
50mM Mg + 50mM Na	65.6°C
100mM Mg + 50mM Na	67.0°C
100mM Na	56.8°C

Looking at the data in Table 8, we see that *Bp*HKA has a very different thermal behavior than *Sw*HKA (Table 6) in the presence of the same concentrations of $MgCl_2$ and HPA. With *Sw*HKA, the highest T_m value was reached when the protein was incubated with magnesium only, decreasing when HPA was added. This is not the case with *Bp*HKA. In the *apo* state, *Bp*HKA has an initial T_m of 55.8ºC, a contrast to *Sw*HKA's T_m of 66.2ºC. When magnesium is added, there's an increase of 6ºC in *Bp*HKA's T_m value and a second increase of about 10ºC after the addition of HPA, for a total T_m shift of 16.4ºC. From these results, the combined presence of magnesium and HPA seems to stabilize *Bp*HKA even more than the presence of magnesium alone.

Longer incubation times of the protein with magnesium did not seem to make a difference in the T_m value, which remained constant at 24 and 48h of incubation, in the two magnesium concentrations tested. *Bp*HKA was also incubated with sodium chloride, to test the hypothesis that the protein's stabilizing factor was the ionic force and not necessarily the presence of the magnesium/metal. As we can see in the last three rows of Table 8, incubation with NaCl does not cause a significant increase in T_m, even when magnesium is also present. This lets us conclude that what leads to the increase of *Bp*HKA thermal stability is the presence of magnesium.

Similarly to what was done with *Sw*HKA, *Bp*HKA was also subjected to a more extensive concentration range of magnesium, HPA and the three aldehydes, DL-glyceraldehyde, glycolaldehyde and butyraldehyde (Table 9).

Table 9 - *Melting temperatures (T_m) obtained through nDSF for BpHKA-WT when incubated with 0 to 1.9M of $MgCl_2$, 0 to 275mM of HPA and 3 different aldehydes.*

Condition		T_m
0mM	Mg	55.8°C
10mM		58.3°C
25mM		59.9°C
50mM		64.0°C
75mM		66.7°C
100mM		67.9°C
150mM		71.0°C
200mM		72.7°C
250mM		74.0°C
300mM		75.1°C
400mM		76.7°C
500mM		77.6°C
600mM		78.8°C
700mM		79.1°C
800mM		79.7°C
900mM		80.0°C
1M		80.2°C
1.1M		80.6°C
1.5M		80.8°C
1.75M		80.6°C
1.9M		79.8°C
0mM	HPA (+200mM Mg)	72.7°C
10mM		73.7°C
25mM		74.4°C
50mM		75.3°C
75mM		76.0°C
100mM		76.7°C
150mM		77.7°C
200mM		78.4°C
250mM		78.6°C
275mM		79.6°C

Condition	T_m
100mM Mg + 100mM HPA + 100mM DL-Glyceraldehyde	75.1°C
100mM Mg + 100mM HPA + 100mM Glycolaldehyde	69.3°C
100mM Mg + 100mM HPA + 100mM Butyraldehyde	72.9°C

Initially, *Bp*HKA was incubated with 0 to 500mM $MgCl_2$, just like *Sw*HKA, but as we can see in Table 9, at 500mM, a defined T_m plateau had not yet been reached, indicating that *Bp*HKA could still be incubated with higher concentrations of $MgCl_2$. The magnesium concentration was then increased gradually to a maximum of 1.9M, which yielded a T_m of 79.8ºC. The highest T_m registered was 80.8ºC, at [$MgCl_2$] = 1.5M, but the plateau of maximum temperatures reached could be considered from 800mM to 1.9M of $MgCl_2$, where all T_m values round 80ºC. When compared to the *apo* state (T_m = 55.8ºC), the presence of magnesium caused an overall increase of about 25ºC in the protein's T_m, reflecting a very accentuated increase in protein stability.

When incubated with both magnesium and hydroxypyruvate (HPA), *Bp*HKA's T_m steadily increased until the maximum value of 79,6ºC, at the highest HPA concentration tested (275mM). Overall, the presence of HPA caused an increase of 6.9ºC in the T_m, with the highest value registered (T_m = 79.6ºC) in the same range as the highest T_m values obtained with only magnesium (T_m ~ 80ºC). However, compared to the 200mM $MgCl_2$ and 275mM HPA condition that yield this high T_m, when incubated with just magnesium, much higher concentrations are needed to reach a T_m ~ 80ºC (800mM-1.9M $MgCl_2$). These results seem to indicate that unlike what happened with *Sw*HKA, HPA has a cumulative stabilization effect on *Bp*HKA, in the presence of magnesium. No incubation was done in the presence of only HPA because without the metal this class of proteins is not able to bind the substrate (HPA) in the active site, or any other place, so no T_m shift would be expected.

Incubation with the different aldehydes showed once again a difference between *Bp*HKA and *Sw*HKA's thermal behavior. Unlike *Sw*HKA, which didn't seem to show a preference between the aldehydes tested, *Bp*HKA yielded very different T_m values depending on the aldehyde. The highest T_m (75.1ºC) was reached with DL-glyceraldehyde, to which *Bp*HKA only binds to the L enantiomer. This was followed by butyraldehyde, with a T_m = 72.9ºC, and the lowest T_m of 69.3ºC,

with glycolaldehyde. Although the focus of this project is the stereoselectivity of *Sw*HKA and *Bp*HKA regarding DL-glyceraldehyde specifically, butyraldehyde and glycolaldehyde were also tested due to their size. Butyraldehyde doesn't have a hydroxyl group but has a longer carbon chain (4C) than DL-glyceraldehyde (3C). Glycolaldehyde has a hydroxyl group and it's the smallest of the three, with a carbon chain of 2 carbon (C) atoms. The different T_m values obtained for the different aldehydes might be due to the size and/or composition of *Bp*HKA's active site that could result in a less stable binding of the other aldehydes. Since nDSF and T_m values should not be used as the sole method in the determination of protein stability and affinity, other experiments should be used to assess both proteins' binding affinity to butyraldehyde and glycolaldehyde. Although not possible in the time frame of this project, the structural determination of *Sw*HKA and *Bp*HKA in complex with butyraldehyde and glycolaldehyde, and comparison with the DL-glyceraldehyde structure, should also prove to be interesting and provide valuable information regarding these protein's reaction mechanism.

3.3. Crystallization screens and trials

3.3.1. *Sw*HKA

The crystallization process of *Sw*HKA mutants S116A and S116C began with a Classic I Suit screen, which led to some crystallization hits. The most promising results yielded thin and triangular-shaped crystals that came from condition G4: 0.2M $MgCl_2$, 0.1M HEPES sodium salt (pH 7.5) and 30% PEG400 (Figure 11).

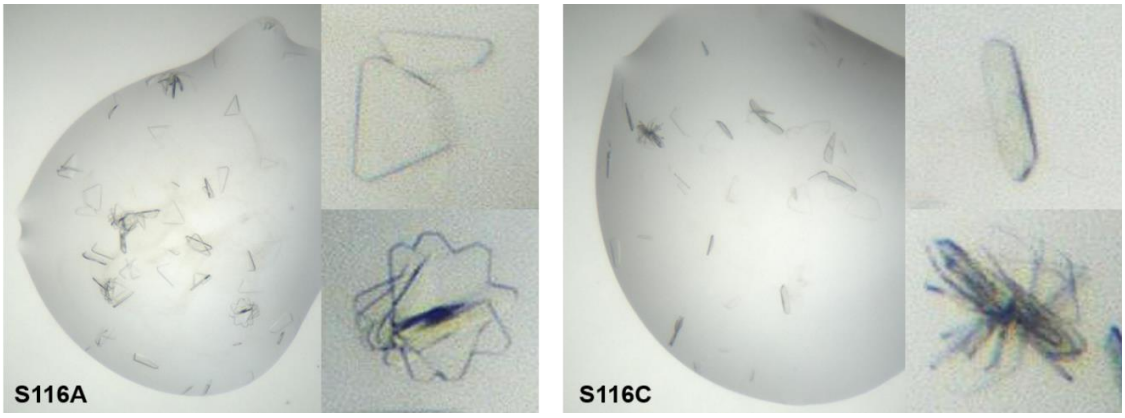


Figure 11 - *Crystallizations hits of S116A and S116C mutants of SwHKA, from position G4 of the Classic I Suit screen.*

The crystals shown in Figure 11 ranged from 50 to 100µm in size and exhibited some inconsistencies regarding crystal shape and multiplicity, with several crystal clusters as well as single crystals. These conditions were then reproduced and optimized manually, by finer concentration screening of the precipitant (PEG400), salt ($MgCl_2$) and buffer pH (HEPES). Some of the resulting drops of this process are shown in Figure 12.

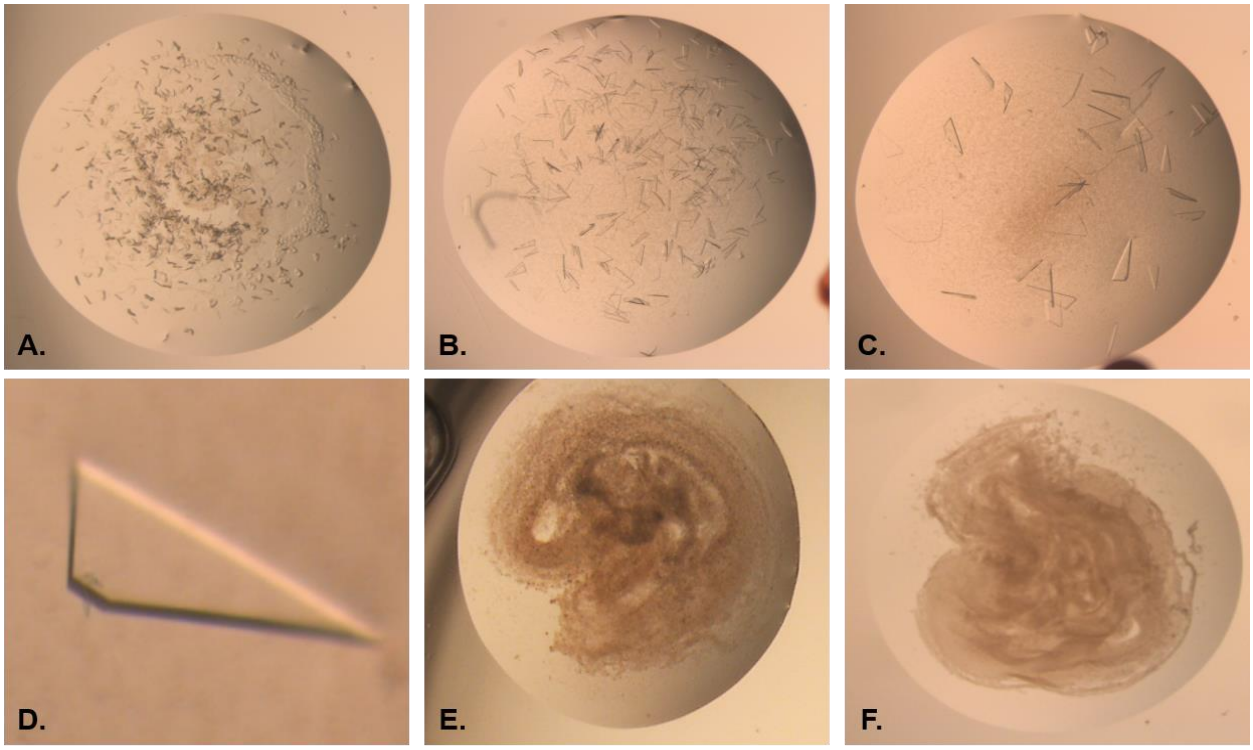


Figure 12 - *Drops of SwHKA-S116A (A to D) and S116C-SwHKA (E and F). Drop conditions: A and E - 0.2M $MgCl_2$, 0.1M HEPES (pH 7), 25% PEG400; B - 0.1M $MgCl_2$, 0.1M HEPES pH 7.5, 20% PEG400; C and F - 0.1M $MgCl_2$, 0.1M HEPES pH 8, 20% PEG400. D – crystal close-up of C.*

Between the two *Sw*HKA mutants, the crystallization process of S116A proved to be the most successful. After the test of several crystallization conditions, a trend became apparent: higher $MgCl_2$ and PEG400 concentrations yielded more nucleation and higher pH caused less nucleation. Single crystals of S116A were fished, soaked and exposed to the X-ray beam. High resolution X-ray data sets (1.4 to 1.8Å) were collected from S116A crystals (Figure 12).

Unfortunately, the crystallization trials of S116C always resulted in nucleation and precipitation (Figure 12, E and F). This lack of crystal growth for the S116C mutant, when compared to the S116A mutant, which crystallized and diffracted well, may be justified by the high degree of degradation of the protein sample, visible in the SDS-PAGE gel (Figure 8). The

crystallization trials of S116C were dropped and the work was focused on the structure solution and refinement of S116A.

In order to obtain the structures of SwHKA-WT in the *apo* (no cofactor present) and *holo*-"all" (cofactor and both substrates present) states, a previously established *Sw*HKA-WT crystallization protocol was replicated. These conditions, consisting of sodium citrate and HEPES buffer, consistently yielded bipyramidal crystals (Figure 13). Crystal size varied depending on the sodium citrate concentration but, on average, the fished crystals had 150-200µm.

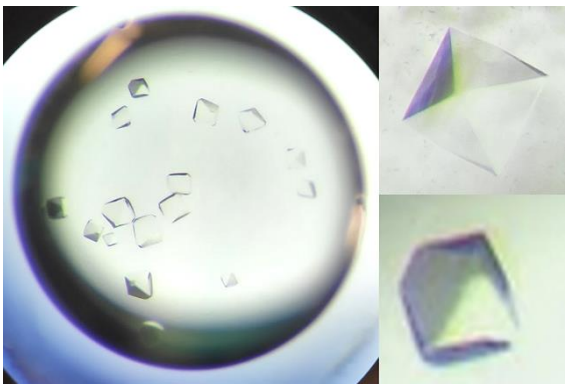


Figure 13 - *Drop of SwHKA-WT in citrate conditions.*

*Sw*HKA crystals always showed high resolution X-ray diffraction, regardless of the age of the crystal. However, if the crystal was exposed to the X-ray beam as soon as they grew, the diffraction limit would only go to *ca.* 2Å or lower. When the crystals were left to age for 2-3 weeks, they were able to diffract the X-rays consistently to around 1.4-1.5Å resolution, and sometimes even going to 1.2Å resolution. Since the crystallization conditions did not have magnesium, the apo-SwHKA structure was determined from a crystal fished directly after the citrate/KBr mother liquid exchange, with no added soaking. Obtaining a structure of *Sw*HKA-all proved to be much more challenging. Although the magnesium ion and the HPA molecule could be seen in the protein's active site at full occupancy after a few minutes of soaking, the presence of D-glyceraldehyde in the structure was much more elusive. After several soaking experiments, varying soaking times from 5 minutes to 24h, DL-glyceraldehyde batch solutions and concentrations (0 to 500mM), the aldehyde began to be visible near the active site at low occupancies. Eventually, the catalyzed product of HPA and D-glyceraldehyde, 2-keto-D-gluconic acid (8YV), still bound to the magnesium ion, was visible in the active site.

3.3.2. *Bp*HKA

Regarding the crystallization process of *Bp*HKA, it began with the manual optimization of previous hit conditions (obtained prior to this project), that consisted of 20% PEG1500, 0.1M bis-tris propane buffer (pH 8.5) and 0.1M $MgCl_2$. The optimization of these conditions focused mainly on the screening of different PEG1500 concentrations but also entailed testing of different drops ratios, crystal growth temperatures and buffer pH values. The best *Bp*HKA crystal drops with PEG1500 as the precipitant are represented in Figure 14.

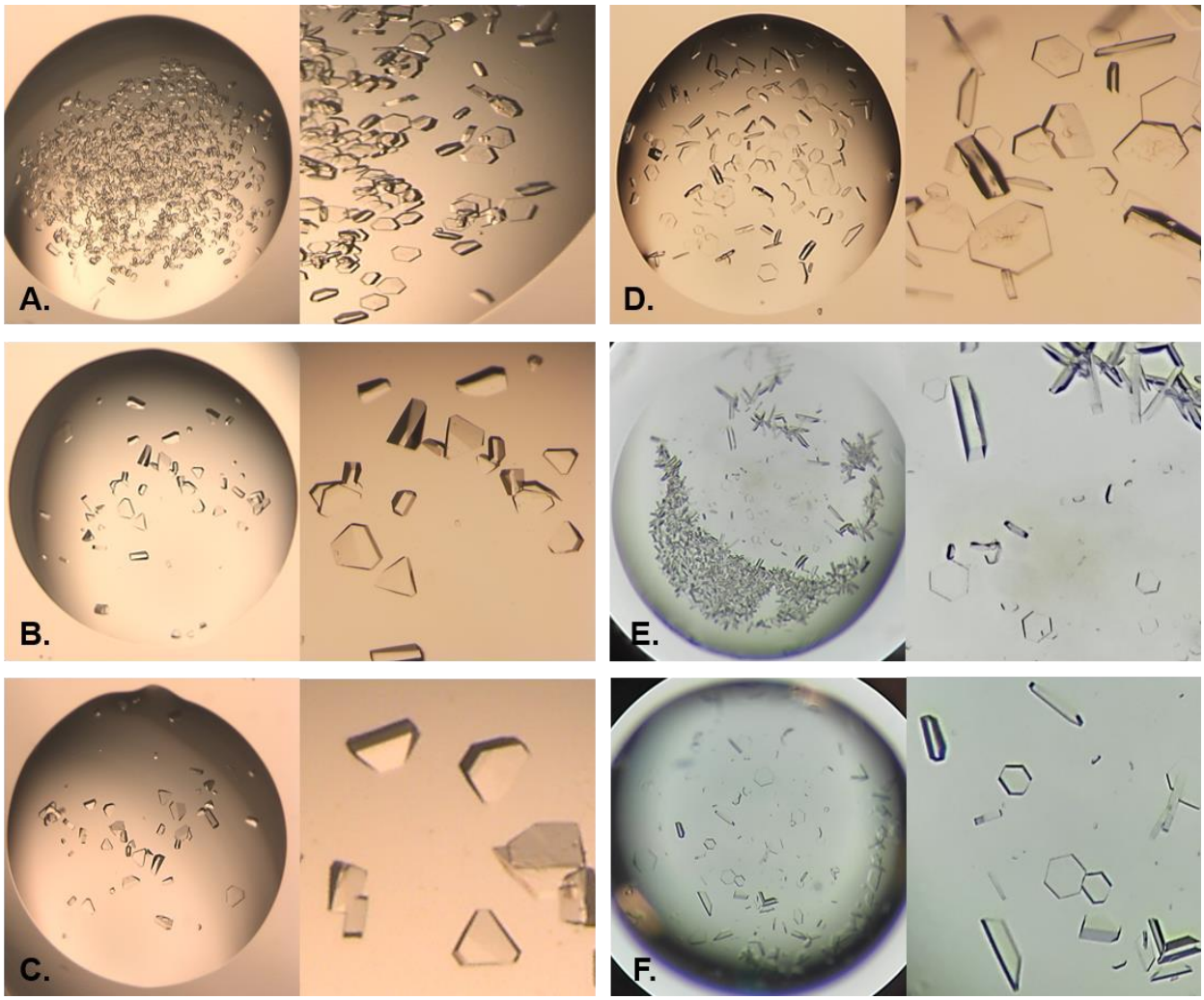


Figure 14 - *Drops of BpHKA with PEG1500 as the precipitant. Reservoir solution (RS) also consisted of 0.1M bis-tris propane at pH 8.5 and 0.1M $MgCl_2$, at 19ºC (A to D) and 4ºC (E and F). Drop ratio of 1.8µL:1.2µL (P:RS) with protein solution at 20.8mg/mL (A), 10mg/mL (B to E) and 8mg/mL (F). Percentage (v/v) of PEG1500: A and B– 20%; C – 18%; D to F – 16%.*

Using PEG1500 as the precipitant agent yielded hexagonal crystals that varied in thickness and size but were usually between 10-15µm and 50-100µm, respectively. The hexagonal shape also varied between perfect hexagons, with six same-sized sides (Figure 14, F)

and imperfect hexagons, with three longer sides and three shorter sides (Figure 14, C). Upon closer look, even though the crystals grew with sharp edges and well-defined surfaces, they tended to have several cracks in the center, as visible in the close-up image of drop D (Figure 14).

Upon X-ray exposure, the crystals diffracted between 1.6 and 2.6Å resolution, but the diffraction patterns showed high degree of multiplicity, probably caused by the cracks in the crystals. Multiplicity is a result of disordered internal arrangement of the crystalline lattice, visible in the diffraction pattern by the imperfect superposition of diffraction spots of symmetry-equivalent reflections, resulting in multiple observable peaks of the same symmetry-equivalent reflection.[15] High degrees of multiplicity can prevent the determination of the correct crystalline symmetry, consequently making it impossible to find the correct structure solution. Crystal growth was also tested at 4ºC, in hopes that the lower temperature would slow down the crystallization process and promote a more ordered internal arrangement of the crystal lattice, decreasing multiplicity. Unfortunately, this did not improve quality of the overall diffraction of the *Bp*HKA crystals.

As an additional effort to optimize the crystallization conditions with PEG1500 as the precipitant, several 96-well CrystalDirect™ plates were set up with custom conditions ranging from 5 to 24% PEG1500, 0.1M bis-tris propane (pH 7 to 8.5) and [*Bp*HKA] = 11, 9 and 5 mg/mL. These drops didn't follow any rational crystal growth-crystallization conditions relation that we could determine, as most drops resulted in precipitation, sometimes with crystal growth among the precipitant. Changing the crystallization/fishing set-up from manual to automated, with the use of the Harvester, didn't make a difference either, as the few crystals that grew did not diffract when exposed to the X-ray beam. Later in this project's timeframe, we discovered the *Bp*HKA degradation problem mentioned previously in the "SDS-PAGE" section, which may be the reason for these confusing crystallization results.

However, at the time, there was no reason to believe that the protein sample could be degraded and the issue was assumed to be the crystallization conditions used. So, as a new approach to the crystallization of *Bp*HKA, the literature was searched for proteins with high sequence identity to *Bp*HKA, whose structure was known and had been determined by X-ray crystallography. The most promising match was a HpcH/HpaI aldolase/citrate lyase family protein (4MF4 PDB ID) from *Burkholderia cenocepacia* J2315, with 75% sequence identity to *Bp*HKA, whose crystallization was done at 15ºC, with a drop ratio of 1.5µL:1.5µL (P:RS) and reservoir solution (RS) of 20% PEG3350 and 0.2M ammonium chloride (AmCl).

With *Bp*HKA's need for a metal cofactor in mind and the available facilities at the lab, 4MF4's conditions were adapted to *Bp*HKA by exchanging AmCl for $MgCl_2$ and changing the temperature from 15ºC to 19ºC. The adapted conditions yielded crystalline precipitant, with a very high quantity of thin triangular microcrystals. After this initial hit, a round of optimization began by lowering the concentrations of the precipitant, PEG3350 (Figure 15).

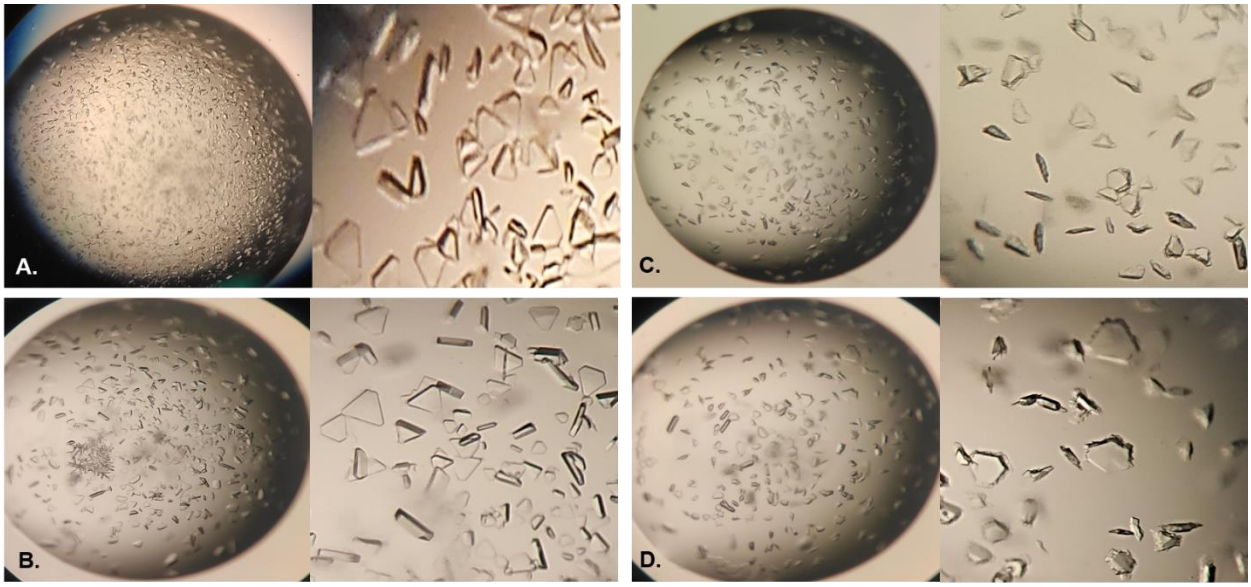


Figure 15 - *Drop of BpHKA with PEG3350 as the precipitant. Reservoir solution (RS) also consisted of 0.2M $MgCl_2$, at 19ºC. Drop ratio of 1.5µL:1.5µL (P:RS) with protein solution at 8mg/mL (A to C), 10mg/mL (D). Percentage (v/v) of PEG3350: A– 14%; B – 12%; C and D – 10%.*

In Figure 15 we can see that lowering the PEG3350 concentration from the initial 20% shifted the system from the precipitation zone to the nucleation and metastable zone of the phase diagram, resulting in crystal growth (section 8.1.1. Crystallization, Figure 1). However, going from 14% to 10% PEG3350 (Figure 15, A and D) did not reduce nucleation significantly. Also, if we look at the close-up image of the drop at 12% PEG3350 (Figure 15, B) and compare it to the close-up of the drops at 10% PEG3350 (Figure 15, C and D), we see that although there's fewer crystals in drops C and D, the crystals seem to have lost their sharp edges and well-defined shape.

To determine if the loss of crystal definition was caused by the low precipitant concentration and if we could further reduce nucleation, drops were set at even lower concentrations of PEG3350 (4-8%). Unlike what happened with the drops shown in Figure 15 (14-10% PEG3350), where crystals were visible after 5 days, drops set with 4-8% PEG3350 were still clear 10 days after the set-up. Due to the COVID-19 pandemic and its consequent restrictions, regular inspection of the drops was not possible for a few weeks after the 10-day mark mentioned.

Two and a half months after the drops were set, an inspection of the drops revealed 2 rod-like crystals in the drop at 8% PEG3350 (Figure 16).

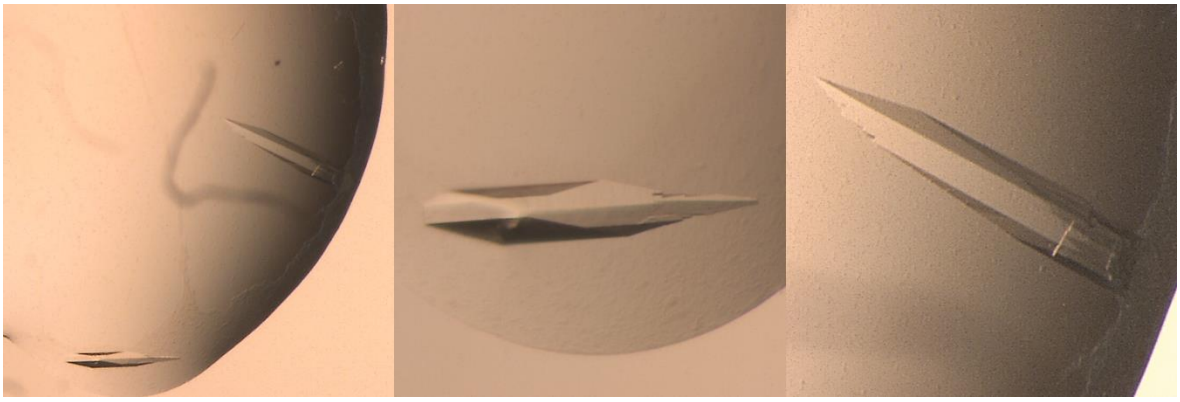


Figure 16 - *Drop of BpHKA with 8% PEG3350 and 0.2M $MgCl_2$, at 19ºC. Drop ratio of 1.5µL:1.5µL (P:RS) with protein solution at 8mg/mL.*

Besides the different shape, the crystals in Figure 16 were 200-300µm long, much bigger than the previous crystals with PEG3350 as the precipitant (Figure 15). During the fishing process, the crystals started to crack and dissolve when touched with the loop, but we were able to fish almost intact crystal pieces. The crystals were then exposed to the X-ray beam, diffracting at around 2-2.5Å resolution.

Although we cannot be sure of exactly how long these crystals took to appear, or if their size and overall shape changed overtime, we know that they took longer than the crystals that grew at higher PEG3350 concentrations. This makes sense because at lower precipitant concentrations, the equilibrium between the concentration of PEG3350 in the drop and in the reservoir/well should take longer to be reached by vapor diffusion. The fact that the reservoir well/coverslip set-up is never 100% air-tight, allowing slow evaporation overtime and changing the equilibrium of the crystallization experiment, may also have had a role in the growth of the crystals. Unfortunately, a new attempt to optimize these last tested conditions and reduce the time that the crystals took to grow was not successful. These experiments consisted of testing different concentrations of $MgCl_2$, protein and PEG3350. However, the crystals represented in Figure 16 were not reproducible and the search for new crystallization conditions began.

Even though the use of PEG1500 and PEG3350 ultimately did not yield good diffraction quality crystals, polyethylene glycol (PEG) did seem to be a promising precipitant agent in the crystallization process of *Bp*HKA. For this reason, a 15-well plate was set with 0 - 0.2M $MgCl_2$ and different PEG polymers (PEG400, PEG500MME, PEG2000MME, PEG4000 and PEG5000MME). After 4 days, several drops had yielded crystals. The best hits are represented on Figure 17.

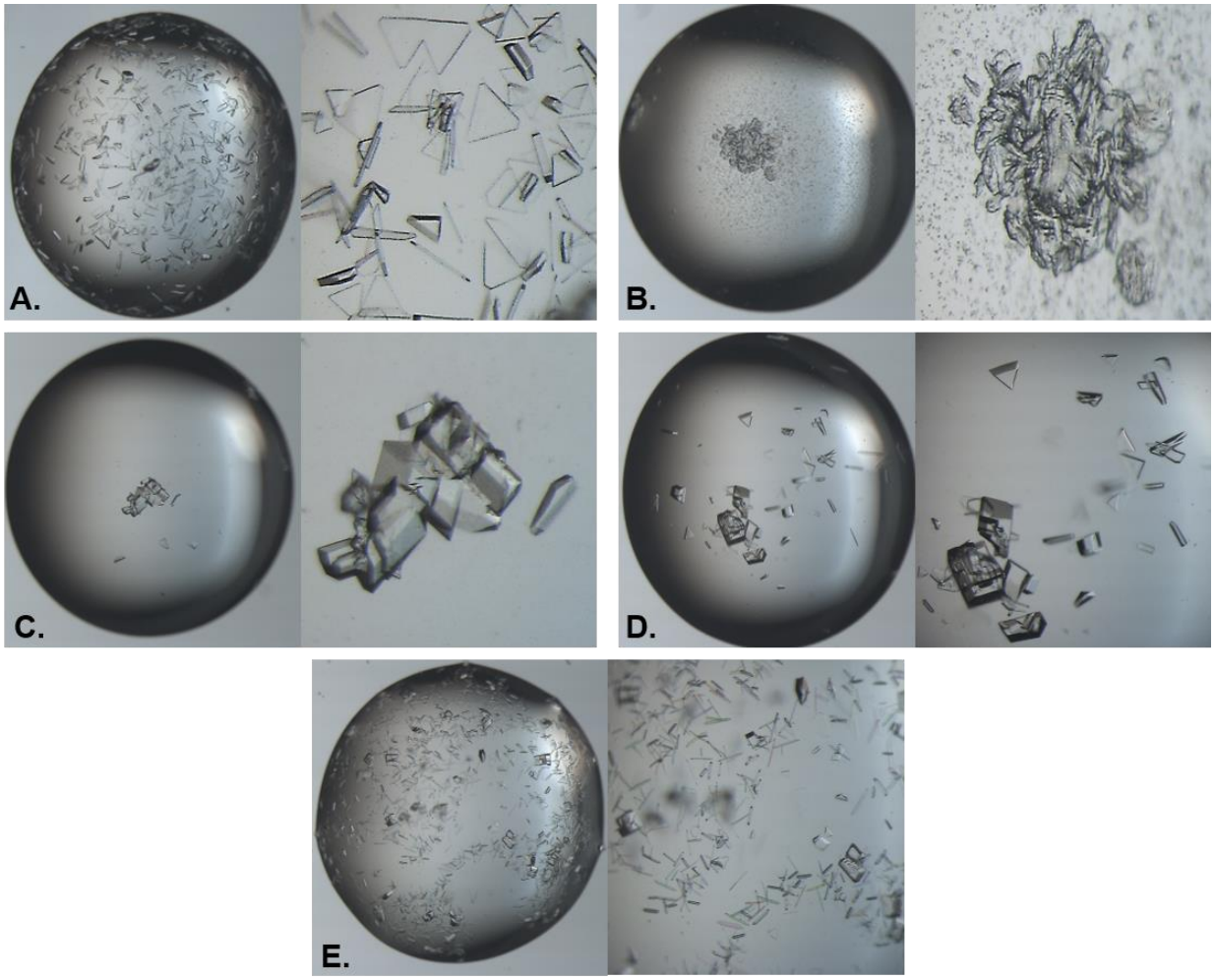


Figure 17 - *Drop of BpHKA with different PEGs as the precipitant. Drop ratio of 1.5µL:1.5µL (P:RS), at 19ºC, with protein solution at 9.5mg/mL. Reservoir solution (RS): A - 25% PEG500MME and 0.1$MgCl_2$; B - 25% PEG500MME and 0.2$MgCl_2$; C - 8% PEG2000MME and 0.1$MgCl_2$; D - 8% PEG4000 and 0.1$MgCl_2$; E - 8% PEG5000MME and 0.1$MgCl_2$.*

From the three $MgCl_2$ concentrations tested (0, 0.1 and 0.2M), 0.1M proved to be the best one, yielding crystals with all the PEG polymers except PEG400. No crystals ever grew in the drops at 0M of $MgCl_2$. Since the conditions tested consist only of precipitant (PEG) and salt ($MgCl_2$), with no buffer solution, the presence of the magnesium ion seems to be vital in the stabilization (as seen previously) and crystal growth of *Bp*HKA. At 0.2M of $MgCl_2$, the only PEG that yielded crystals after 4 days was PEG500MME (Figure 17, B).

Regarding the different PEG polymers, as expected, each one of them yielded different results. In the drops with PEG400 no crystals ever grew, making PEG400 most unsuccessful of the polymers tested regarding *Bp*HKA's crystallization, even though PEG400 proved to be the best cryoprotectant for *Bp*HKA crystals, regardless of the crystallization condition. As we can see in Figure 17, depending on the PEG polymer the drops yielded: microcrystals with PEG5000MME

(E); triangular crystals with PEG500MME (A), similar to the hexagonal crystals studied before (Figure 15); and crystal clusters with PEG2000MME (C) and PEG4000 (D). Also noteworthy is the fact that with PEG500MME, crystal growth was very different at 0.1 and 0.2M MgCl2 (Figure 17, A and B respectively).

From Figure 17, drops C and D, corresponding to PEG2000MME and PEG4000, were chosen for optimization through seeding experiments. For this, new drops were set with the same crystallization conditions as before (C - 8% PEG2000MME and 0.1$MgCl_2$; D - 8% PEG4000 and 0.1$MgCl_2$) and seeds from the drops shown in Figure 17 were transferred to these new drops. The results of the seeding experiment are shown in Figure 18.

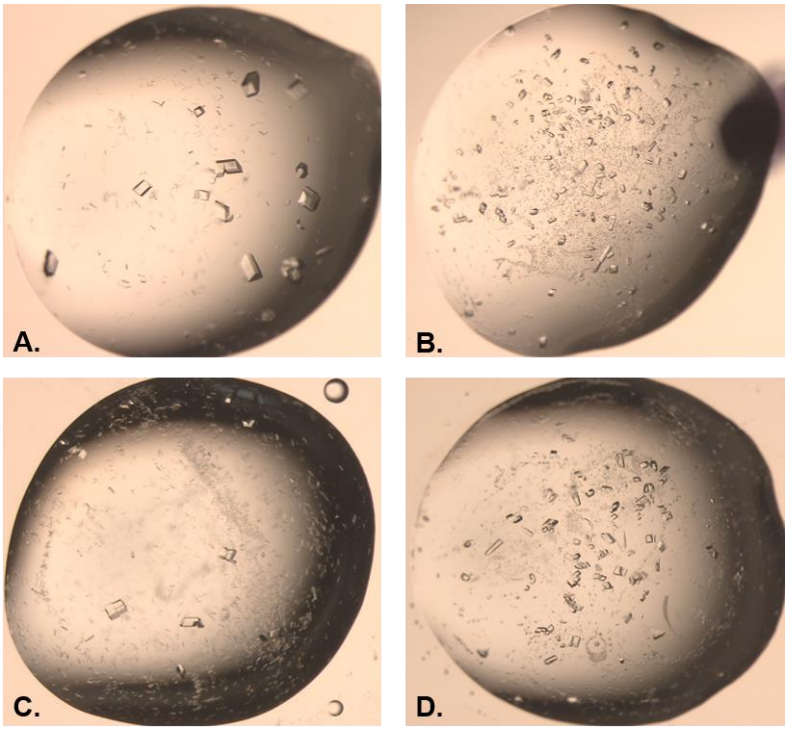


Figure 18 - *Drop of BpHKA with seeding. Reservoir solution (RS) with 8% PEG2000MME (A and B) and 8% PEG4000 (C and D), and 0.1M $MgCl_2$ (all). Drop ratio of 1.5µL:1.5µL (P:RS), at 19ºC, with protein solution at 8mg/mL (A and C) and 7mg/mL (B and D). A and C – no seeding; B and D: seeding with seeds from the crystals in Figure 17.*

Comparing the drops with no seeding (Figure 18, A and C), to the previous drops in the same conditions (Figure 17, C and D), we see that the drops look very different. The drops in Figure 18, set as the ones in Figure 17, did not yield crystal clusters as before, yielding microcrystals and single crystals of around 30µm in size instead. At this point in time, we were already aware of the degradation of *Bp*HKA at 4ºC and the drops were set with freshly defrosted protein sample, excluding that as the reason for the difference in the drops. In the drops where seeding was done, a large number of small crystals grew with some microcrystals as well. The

lack of success of this seeding experiment may come from the possibility that too many seeds of the original crystal clusters, shown in Figure 17, were transferred to the drops.

The seeding process was repeated and the drops were left at 4ºC, hoping that the lower temperature and consequent slower crystallization rate would promote the growth of the crystal seeds instead of the growth of critical nuclei. Unfortunately, at 4ºC the results were the same as at 19ºC (Figure 18). Using different protein and precipitant concentrations did not cause a change in the crystals and their ability to diffract either.

Going forward, the PEG2000MME crystallization conditions (8% PEG2000MME and 0.1$MgCl_2$) were chosen for further optimization since they yielded the best-looking and more individual crystals, when compared to PEG4000 (Figure 18, A and C).

Crystals of the drop shown in Figure 18 (A, no seeding), were fished and exposed to the X-ray beam. The crystals diffracted from 2 to 2.3Å resolution and the diffraction pattern showed high degree of multiplicity.

In order to optimize the PEG2000MME condition (8% PEG2000MME and 0.1$MgCl_2$), different experiments were done simultaneously: setting “Chinese” drops (Figure 19), additive screen (Figure 20) and a dehydration experiment (Figure 21).

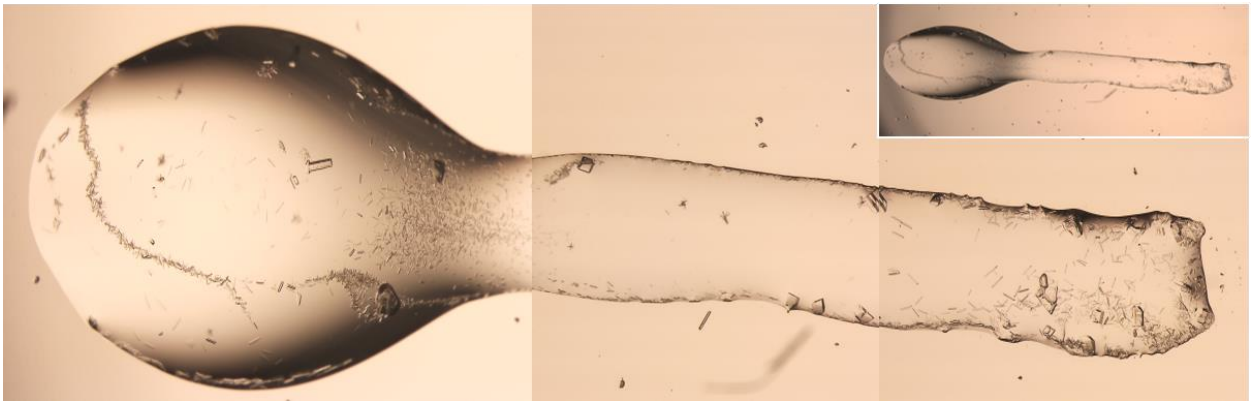


Figure 19 - Drop of BpHKA set with the “Chinese” drop technique. Reservoir solution (RS) with 8% PEG2000MME and 0.1M $MgCl_2$, at 19ºC. Drop ratio of 1.5µL:1.5µL (P:RS) with protein solution at 8mg/mL.

In the Chinese drop technique, the protein solution is not mixed with the reservoir solution/mother liquid like what happens in a “normal” vapor diffusion drop set-up. Instead, the protein sample (P) and the reservoir solution (RS) are only put in contact on top of the cover slip, and the two liquids mix slowly over time through diffusion. Although eventually the drop reaches an equilibrium, this technique adds a dynamic element to the crystallization experiment, forming a P:RS gradient inside the drop itself.

As we can see in Figure 19, the left section of the drop, where the RS was pipetted and its concentration was higher in the initial stages of crystal growth, shows a lot more nucleation and microcrystals when compares to the right side of the drop, the furthest from the RS. The irregular volume distribution of the drop causes the rough edges visible on the right side of the drop due to the low volume of the drop region. The best region for *Bp*HKA crystal growth seems to be in the center of the drop, a middle place between the protein solution and the RS. In this central part of the drop, we find single crystals of roughly 40µm. These crystals were fished and soaked with $MgCl_2$ and HPA. When exposed to the X-ray beam, the crystals diffracted well, between 1.4 and 1.8Å with little to no amount of multiplicity.

Regarding the additive screen, the best hits are shown in Figure 20.

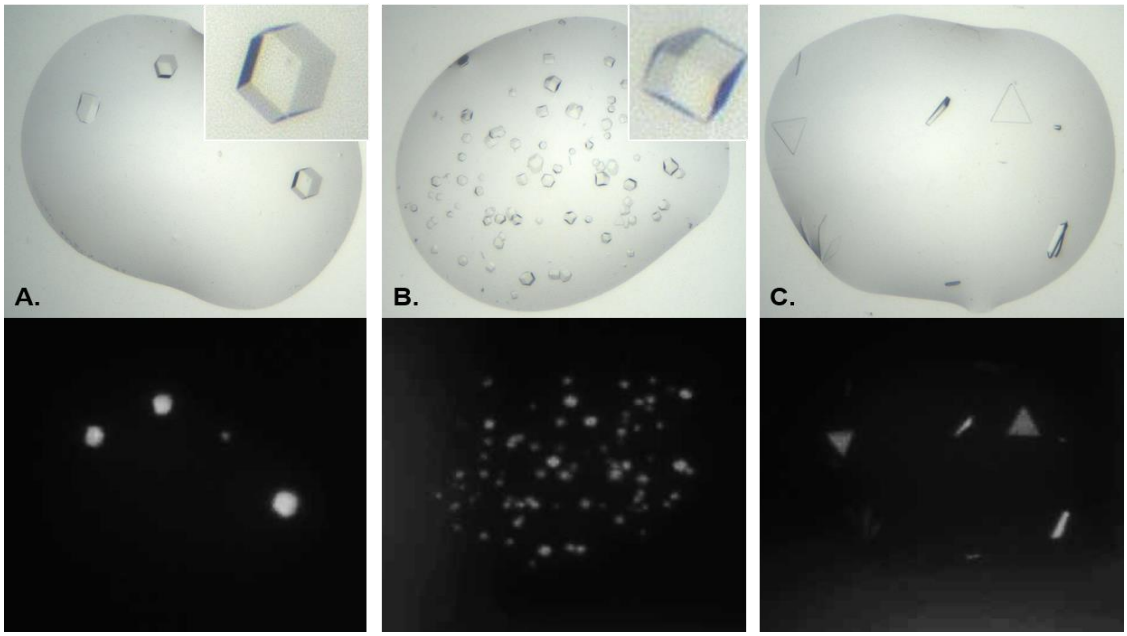


Figure 20 - *Crystallizations hits of BpHKA, from position H5, H7 and F9 of the Additive screen. UV inspection confirms the presence of protein crystals. Additive: A - 4% formamide; B - 0.5% ethyl acetate; C - 2% benzamidine hydrochloride.*

The crystals shown in Figure 20, as well as a few others, were harvested directly from the plate, using the harvester, but due to a Harvester malfunction, some of the crystals were lost. The crystals that were successfully harvested were exposed to the x-ray beam but some didn't diffract at all and others diffracted only at resolutions lower than 3.0Å. The conditions with formamide as an additive were reproduced by hand, but when exposed to the X-ray beam, they showed no diffraction.

Finally, a dehydration experiment was performed with the PEG2000MME crystallization conditions, to see if the dehydration would help crystal packing, decreasing multiplicity. For this, several drops were set at 8% PEG2000MME and 0.1$MgCl_2$. The drops were left to grow for 5 days, after which the coverslips were changed to wells with reservoir solutions at higher PEG2000MME concentration. No pipetting or mixing was done at this point and the coverslips were merely changed to wells with higher concentrations of PEG2000MME (and the same concentration of $MgCl_2$) and then sealed again (Figure 21). These conditions were inspired by *Sw*HKA, whose diffraction improves with aging/dehydration of the crystal drops. Because we were almost at the end of the timeframe of this project, the cover-slip exchange to higher PEG2000MME concentrations was done as a faster simulation of the effect of the slow evaporation of the crystallization conditions.

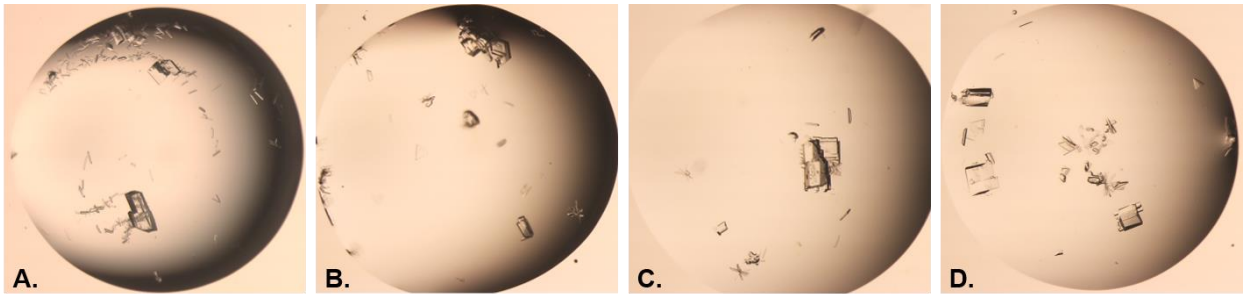


Figure 21 - *Drop of BpHKA after a forced dehydration experiment. All drops were set with reservoir solution (RS) of 8% PEG2000MME and 0.1M $MgCl_2$, at 19ºC, drop ratio of 1.5µL:1.5µL (P:RS) and protein solution at 8mg/mL. New PEG2000MME concentration: A – control, no well change; B – 12%; C – 16%; D – 20%.*

Five days after the well exchange, the crystals looked the same as before, with no visible change in their size or shape. When exposed to the x-ray beam, the diffraction did not improve when compared to the “original” crystals (Figure 18, A) and kept showing high degrees of multiplicity.

At the end of this project, during a routine inspection of all the plates set during the length of the project, a new crystal was found in the condition set with 8%PEG5000MME and 0.2M $MgCl_2$, 3 months after the drops were set (Figure 22). This condition belonged to the screening of different PEG polymers shown in Figure 17, and the plate was inspected regularly for the first 6 weeks. During those inspections, they stayed clear, with no sign of crystallization.

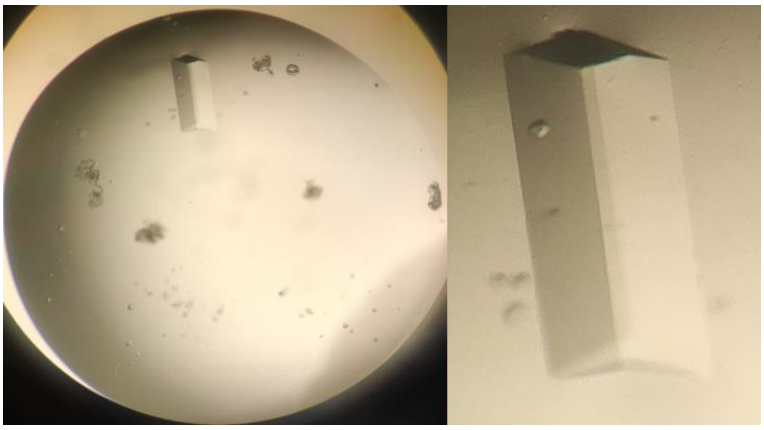


Figure 22 - *Drop of BpHKA with 8% PEG5000MME and 0.2M $MgCl_2$, at 19ºC. Drop ratio of 1.5µL:1.5µL (P:RS) with protein solution at 9.5mg/mL.*

Figure 22 shows a thick rod-like crystal, with rectangular sides. The crystal was 500-600µm long, much bigger than all other *Bp*HKA crystals yielded during this project. Due to its large size, the crystal was broken into smaller pieces of around 80-100µm, with the help of crystallization tools. The smaller crystal pieces were then fished and soaked with all 3 substrates of interest (Mg^{2+}, HPA and L-glyceraldehyde), for different periods of time. The crystal pieces diffracted between 1.3-1.7Å resolution, with super clean diffraction patterns of intense spots and no signs of multiplicity.

This discovery happened in the last week of the project so there was no time to optimize or replicate the conditions with PEG5000MME. However, these conditions would be a great starting point for future *Bp*HKA crystallization projects.

3.4. Overall crystal structures

During this project a total of 5 different structures were determined: 2 structures of *Sw*HKA-WT, 2 structures of *Sw*HKA mutant S116A and 1 structure of *Bp*HKA. No X-ray diffraction data was ever collected for the S116C mutant.

3.4.1. *Sw*HKA

Apo-*Sw*HKA and *Sw*HKA-all, determined from the crystals shown in Figure 13, crystallized with one molecule in the asymmetric unit, in space group $P4_232$ and unit cell parameters a,b,c = 116.38Å and a,b,c = 115.85Å, respectively, and $\alpha,\beta,\gamma = 90º$.

S116A-Mg crystallized in the cubic space group H3, with unit cell parameters a,b = 71.006Å, c = 222.48Å, α,β = 90º and γ = 120º, and two molecules in the asymmetric unit. The crystal used for the soaking is shown in Figure 12, D.

The S116A-Mg-HPA structure was determined with a crystal from the same drop as S116A-Mg, and also crystallized with two molecules in the asymmetric unit and space group H3 but had unit cell parameters of a,b = 71.327Å and c = 223.19Å.

The overall data collection and refinement statistics of the four *Sw*HKA structures determined by X-ray crystallography in this project is shown in Table 10.

Table 10 - *Data collection and refinement statistics of SwHKA-WT, structures apo-SwHKA and SwHKA-all, and SwHKA-S116A, structures S116A-Mg and S116A-Mg-HPA. Parameters with an asterisk "*" correspond to outer shell values.*

		***apo-Sw*HKA**	***Sw*HKA-all**	**S116A-Mg**	**S116A-Mg-HPA**
Space group		P 42 3 2	P 42 3 2	H 3	H 3
Resolution (Å)		52.04 - 1.35 (1.37 - 1.35)*	81.92 - 1.80 (1.84 – 1.80)*	74.16 - 1.70 (1.73 - 1.70)*	74.39 - 1.80 (1.84 - 1.80)*
Data completeness (%)		100 (100)*	99.9 (100)*	100 (100)*	98.5 (99.0)*
Multiplicity		20.3 (14)*	20.9 (20.9)*	10.4 (10.5)*	3.4 (3.2)*
$<I/\sigma(I)>$		12.9 (1.7)*	30.4 (3.8)*	12.5 (1.8)*	8.5 (1.2)*
$CC_{1/2}$		0.999 (0.723)*	0.999 (0.923)*	0.998 (0.658)*	0.997 (0.365)*
R-merge		0.11 (2.319)*	0.056 (0.913)*	0.130 (1.329)*	0.109 (1.190)*
R_{meas}		0.114 (2.405)*	0.057 (0.936)*	0.137 (1.398)*	0.127 (1.407)*
R_{pim}		0.025 (0.627)*	0.013 (0.204)*	0.042 (0.429)*	0.064 (0.0737)*
Total reflections		1204977	527197	478502	129868
Unique reflections		59220	25246	46039	38705
R		0.132	0.190	0.224	0.183
R_{free}		0.161	0.226	0.247	0.225
Number of non-hydrogen atoms	macromolecules	1961	1881	3825	3802
	ligands	25	36	10	9
	solvent	331	162	567	299

	***apo-Sw*HKA**	***Sw*HKA-all**	**S116A-Mg**	**S116A-Mg-HPA**
Protein residues	253	251	503	499
RMS(bonds) (Å)	0.47	0.60	0.87	0.81
RMS(angles) (º)	0.72	0.97	1.0	0.94
Ramachandran favored (%)	262 (98%)	238 (95%)	497 (97%)	487 (97%)
Ramachandran allowed (%)	5 (2%)	9 (4%)	13 (2%)	12 (2%)
Ramachandran outliers (%)	1 (0%)	3 (1%)	2 (0%)	3 (1%)
Rotamer outliers (%)	3 (2%)	4 (2%)	3 (1%)	6 (2%)

Some “rules of thumb” that helped guide the data solution and refinement process were that a R value is considered good when it reached a value 10 times lower than the high-resolution limit of the data and that the difference between R and R_{free} should not exceed 5%.[17] The R and R_{free} decreased through several rounds of refinement cycles and model building where residues, ligands and water molecules were added to the structure, ensuring that there were no major clashes or bond strains. Ramachandran and rotamer outliers were corrected when possible.

In the specific case of the S116A-Mg structure, the R and R_{free} values are a bit elevated and do not follow the “rule of thumb” mentioned above. However, the refinement of this structure converged to the values specified in Table 10, and the obtained structure makes biological, geometrical and chemical sense, and does not show unjustified electron densities blobs in the electron density maps. The higher values of R and R_{free} may come from the high redundancy/multiplicity of the data (10.4 overall and 10.5 in the outer shell), and from the presence of translational non-crystallographic symmetry in this data set.

During the rounds or refinement cycles, determining the occupancy of the ligands was also possible. The occupancy reflects the presence of a molecule in a specific spot of a three-dimensional structure.[15] If a molecule has full occupancy (occ = 1), it means that the molecule and/or atom is in that location 100% of the time. Lower occupancies usually reflect a mobile molecule. A schematic overview of Mg and HPA’ occupancies observed for every *Sw*HKA structure described in this project is shown in Figure 23.

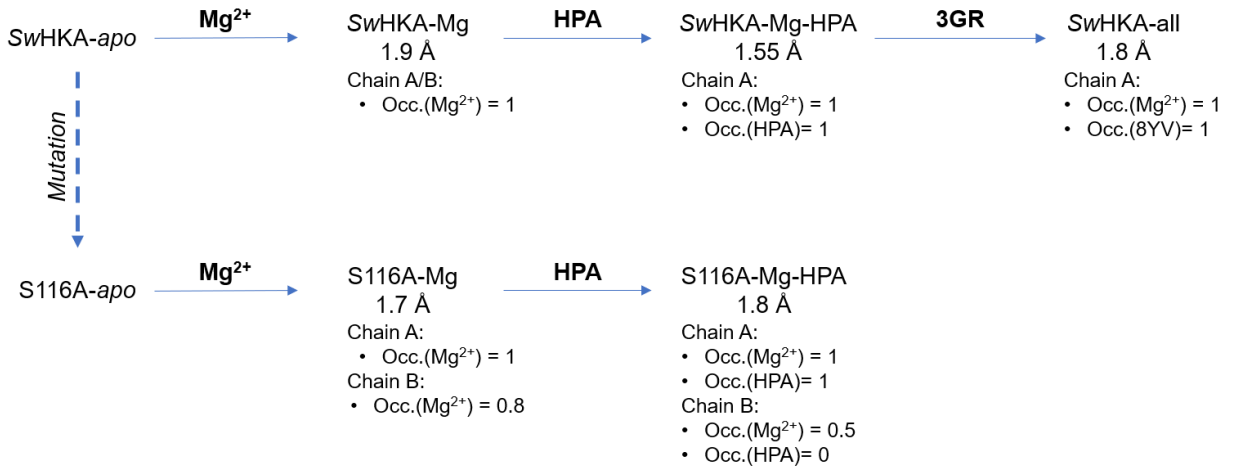


Figure 23 - *Schematic overview of the refined SwHKA structures. HPA – hydroxypyruvate; 3GR – D-glyceraldehyde; 8YV – 2-keto-D-gluconic acid (product of the aldol reaction catalyzed by SwHKA).*

The structural analysis of *Sw*HKA was done through the study of the protein in the *apo* and *holo* states, with aim of identifying the conformational changes induced by the presence of the cofactor and both substrates of interest, hydroxypyruvate (HPA) and D-glyceraldehyde, in the active site. To do this, the 4 structures of *Sw*HKA-WT were superimposed, and the results are shown in Figure 24. Although only the structures of *apo*-*Sw*HKA and *Sw*HKA-all were determined during this project, the previously determined structures of *Sw*HKA-Mg and *Sw*HKA-Mg-HPA (PDB ID 6R62), were also necessary for the better understanding of the protein behavior in the presence of magnesium and the substrates.

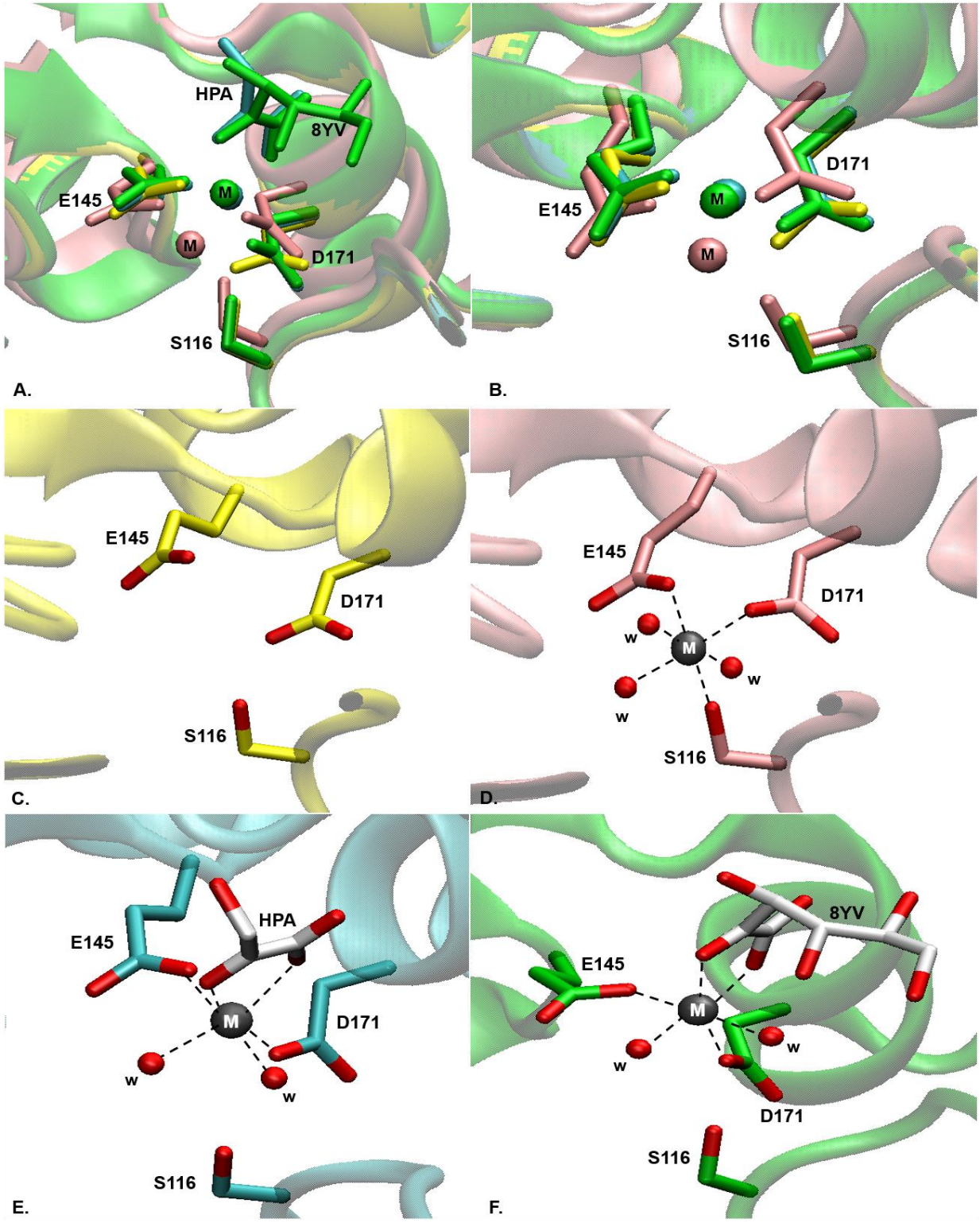


Figure 24 - *Representation of the active site of SwHKA-WT, from different structures: amino acids S116, E145 and D171, water molecules (w), magnesium ion (M), hydroxypyruvate (HPA) and 2-keto-D-gluconic acid (8YV). Magnesium, HPA and 8YV are present in the structures at full occupancy. A - superimposed structures of apo-SwHKA (yellow), SwHKA-Mg (pink), SwHKA-Mg-HPA (cyan) and SwHKA-all (green). B - zoom-in of A, without HPA and 8YV representation. C - apo-SwHKA; D - SwHKA-Mg; E - SwHKA-Mg-HPA; F - SwHKA-all.*

As stated in the Introduction (section 1.2), class II pyruvate aldolases adopt an homo-hexameric conformation in the active form, with a total of six active sites per hexamer. In the specific case of *Sw*HKA, the active site is formed by 2 catalytic amino acids residues, Glu145 and Asp171, as well as a serine residue (Ser116) from a neighboring chain (Figure 24, C). The superposition of all *Sw*HKA-WT structures allows us to analyze the movement of these key residues during the different stages of the aldol reaction. Through the superposition of all the different *Sw*HKA structures we also see that there is no major difference in the fold of any of the structure, indicating that no significant conformational changes happen in the protein during the reaction. In Figure 24, panels A and B, we see that during the different steps of the reaction the side chains of the catalytic residues (Glu145 and Asp171) seem to move to better accommodate the magnesium ion and the substrates.

Throughout the reaction, the coordination sphere of the magnesium ion is completed by different atoms. When only magnesium is present in the active site, represented by the *Sw*HKA-Mg structure (Figure 24, D), magnesium binds to the catalytic residues E145 and D171, 3 water molecules and S116 from a neighboring chain. We call this the resting state, where magnesium is bound to the protein, in a place different from the catalytic one, with no reaction occurring. When HPA enters the active site, represented by the *Sw*HKA-Mg-HPA structure (Figure 24, E), we move to the active state, where the magnesium loses its coordination to the serine residue and to one of the water molecules, and becomes coordinated instead by 2 carbonyls groups of HPA. When DL-glyceraldehyde is added to the system, the protein is stereoselective for the D isomer, which will approach the C3 of the HPA, and *Sw*HKA will catalyze the carbon-carbon bond formation between the C3 from HPA and C1 from D-glyceraldehyde. This will form 2-keto-D-gluconic acid (8YV), where the C4 is equivalent to C1 of the D-glyceraldehyde. This reaction is reversible, and stereoselective regarding the C4. In the *Sw*HKA-all structure, (Figure 24, F), we see the product of the reaction (8YV), which means that at least one reaction cycle had already occurred. In this structure we can also see that the coordination sphere of the magnesium ion stays the same as it was in the SwHKA-Mg-HPA structure, since the HPA oxygen atoms that bind to the magnesium are double-coordinated to C1 and C2, which are not affected by the aldol reaction. Although the elements of the coordination sphere of the magnesium ion change during the reaction, the octahedral geometry and the full occupancy of the metal are kept in both the resting and the active state of the enzyme.

For a better look at the changes that occur in *Sw*HKA's active site during the reaction, the different structures were superimposed sequentially (Figure 25).

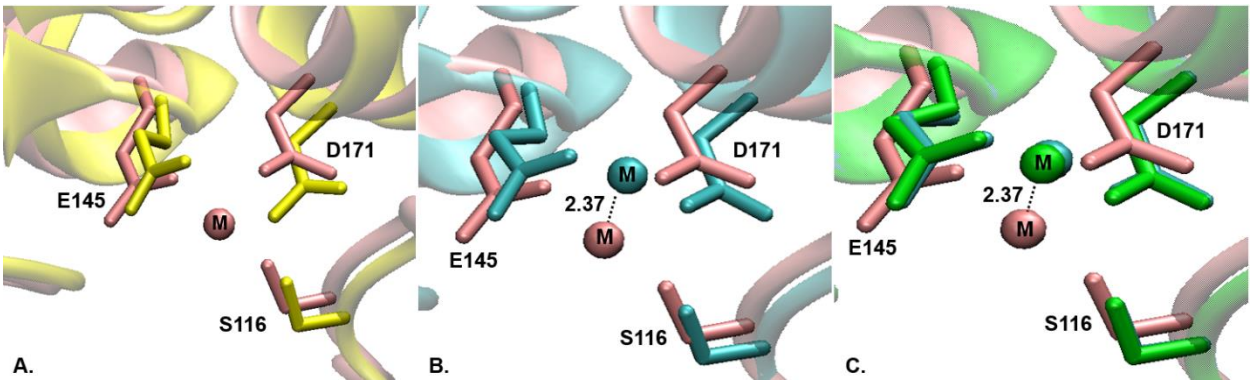


Figure 25 - *Superimposed representation of the active site of different structures of SwHKA-WT. Water molecules not shown for schematic clarity. A - apo-SwHKA (yellow) and SwHKA-Mg (pink); B – SwHKA-Mg (pink) and SwHKA-Mg-HPA (cyan); C - SwHKA-Mg (pink), SwHKA-Mg-HPA (cyan) and SwHKA-all (green).*

In Figure 25, A, we see a shift in the position of the protein backbone and the side chains of the amino acid residues of interest from the *apo* (in yellow) to the *holo* state (in pink). The presence of the magnesium ion causes a conformational change that brings the side chains of S116, E145 and D171 closer to each other, in order to form the coordination sphere of the magnesium ion. When HPA reaches the active site (Figure 25, B), the magnesium ion shifts about 2.4Å, moving out of reach of S116 (structure in pink) and establishing the active state conformation (structure in cyan). As we can see by the overlap in the SwHKA-Mg-HPA (cyan) and SwHKA-all (pink) structures in Figure 25.C, after the initial shift, the magnesium ion stays in the same position during the rest of the reaction.

Because the resting state of the enzyme was isolated and identified for the first time in these structures, we wanted to understand if it was a significant to the enzymatic mechanism, or if it was an artefact of the crystallization and soaking process. Therefore, we decided to access the importance of S116 in the formation of the resting state, so the structure of a *Sw*HKA serine mutant, S116A, in complex with magnesium, was determined (Figure 26).

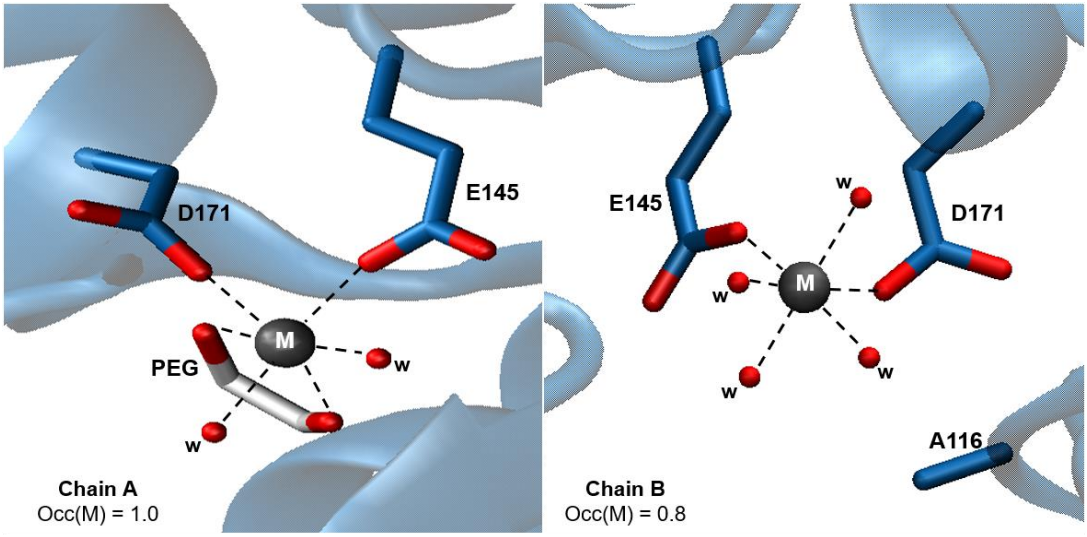


Figure 26 - *Representation of the active site of SwHKA-S116A-Mg.*

Unlike what happened with *Sw*HKA-WT, *Sw*HKA-S116A crystallized with two molecules in the asymmetric unit and space group H3, which means that there are two active site conformations co-existing in the same crystal, at the same time, in different chains (Figure 26). Since the crystallization conditions directly affect the arrangement of the atoms in the crystal lattice, it comes as no surprise that *Sw*HKA-WT and *Sw*HKA-S116A crystallized in different space groups and with different asymmetric units because of the different crystallization conditions used with both proteins.

In chain A, the magnesium ion is present at full occupancy, coordinated to E145, D171, 2 water molecules and a PEG molecule. PEG is part of the crystallization conditions of SwHKA-S116A and its role in the stabilization of the magnesium ion is an artifact of crystallization, an intermolecular contact/interaction that occurs solely as the result of protein crystallization.[45] Since PEG is not a molecule that *Sphingomonas wittichii* synthesizes or consumes, this particular interaction does not occur in nature and is not relevant to our study.

In chain B, the magnesium ion is coordinated to E145, D171 and 4 water molecules, with an occupancy of 80%. Due to the mobile nature of the water molecules in the solvent channels inside the crystals, the magnesium coordination sphere is not as stable as the previous examples (SwHKA-WT-Mg), reflected by the lower occupancy of the metal in the active site.

In both chains, the mutated alanine residue is too far away to interact with the metal, as expected. Alanine was chosen for the serine mutation because of the unlikelihood that it would be able to coordinate and stabilize the magnesium ion due to its methyl side chain, which is too

short to be able to reach the metal site without significant conformational changes of the protein structure. A bond between the magnesium ion and the methyl group (Cβ) could also be considered very unlikely in this case.

To better compare the active sites of the wild type (WT) and the mutant protein, the two structures were superimposed, and the results are shown in Figure 27. The overall fold of both proteins was similar, with no significant differences apart from the active site.

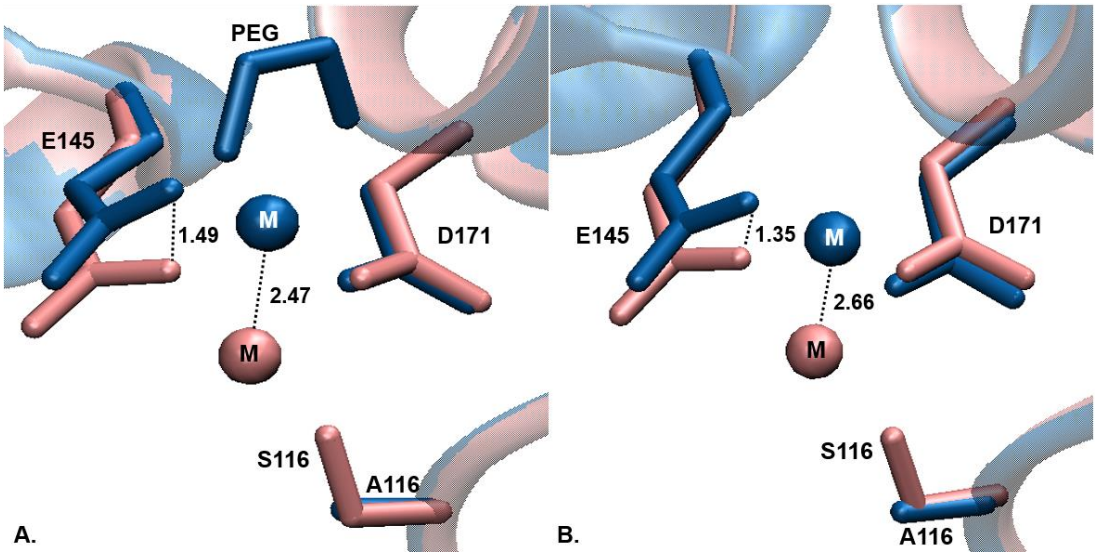


Figure 27 - *Superimposed representation of the active sites of SwHKA-Mg (pink) and S116A-Mg (blue). Water molecules not shown. A – chain A active site of S116A-Mg; B – chain B active site of S116A-Mg.*

The most noticeable difference observed between the *Sw*HKA-Mg (in pink) and S116A-Mg (in blue) structures is the shift in the position of the magnesium ion, caused by the absence of the resting state conformation in the S116A mutant structure. More so, the position of the metal ion in both chain A and chain B is almost identical to the metal position seen previously in the structure of *Sw*HKA-Mg-HPA (Figure 25, B), after the shift to the active conformation. Although the distance of the magnesium ions between the *Sw*HKA-Mg and the S116A-Mg structures (2.47Å and 2.66Å) is bigger than the distance between the *Sw*HKA-Mg and *Sw*HKA-Mg-HPA structures (2.37Å), it seems that without the serine residue, the magnesium does not adopt the resting state conformation and coordinates directly in the active state conformation.

Another difference between the structures is the position of the side chains of the catalytic residues, most noticeably E145. Conformational changes of the side chains of the residues near the active site are to be expected since the proteins must accommodate the cofactor and the substrates, ensuring that the bond lengths are favorable and that there are no major clashes

between the substrates and the protein residues. As glutamate has a longer side chain than aspartate, it makes sense that E145's side chain has to move more than D171's side chain.

To access if the non-existence of the resting state conformation also affected metal occupancy when the substrates were present in the active site, the structure of S116A in complex with magnesium and hydroxypyruvate (S116-Mg-HPA) was determine (Figure 28). Once again, there were two different active site conformations depending on the chain (A and B), and in both, A116 is too far away to coordinate the magnesium cation.

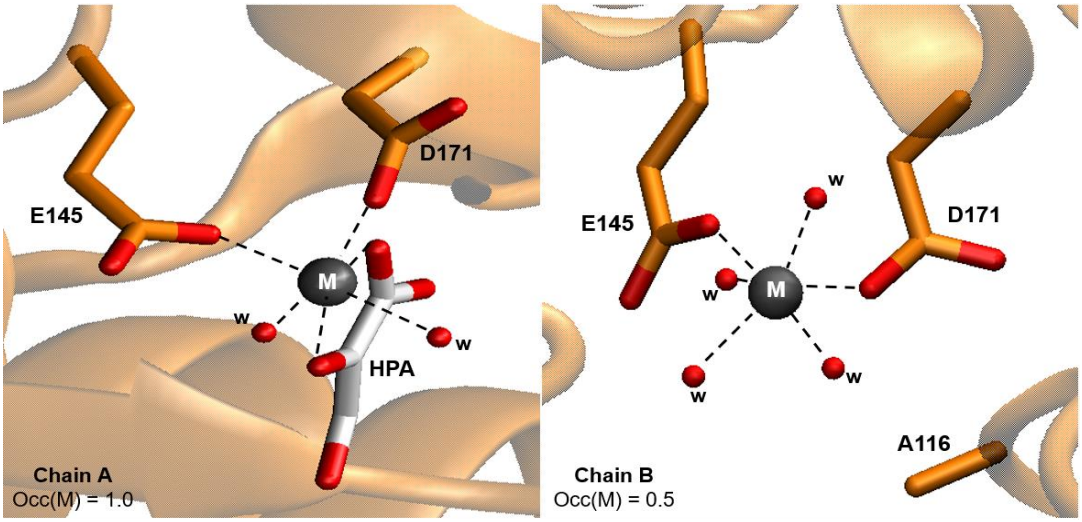


Figure 28 - *Representation of the active site of SwHKA-S116A-Mg-HPA.*

The magnesium coordination sphere in chain A proved to be the same as *Sw*HKA-Mg-HPA, with the metal bound at full occupancy to the two catalytic residues, 2 water molecules and the 2 carbonyl groups of HPA. In this conformation, consistent with the active state, the HPA molecule is also present at full occupancy. In chain B, the magnesium is only 50% present, binding to E145, D171 and 4 water molecules, in a similar way to the B chain of S116A-Mg.

When we compare S116A-Mg-HPA to SwHKA-Mg-HPA (Figure 29), we see that the overall fold of both proteins was similar, with no significant differences apart from the active site, depending of the S116A-Mg-HPA chain.

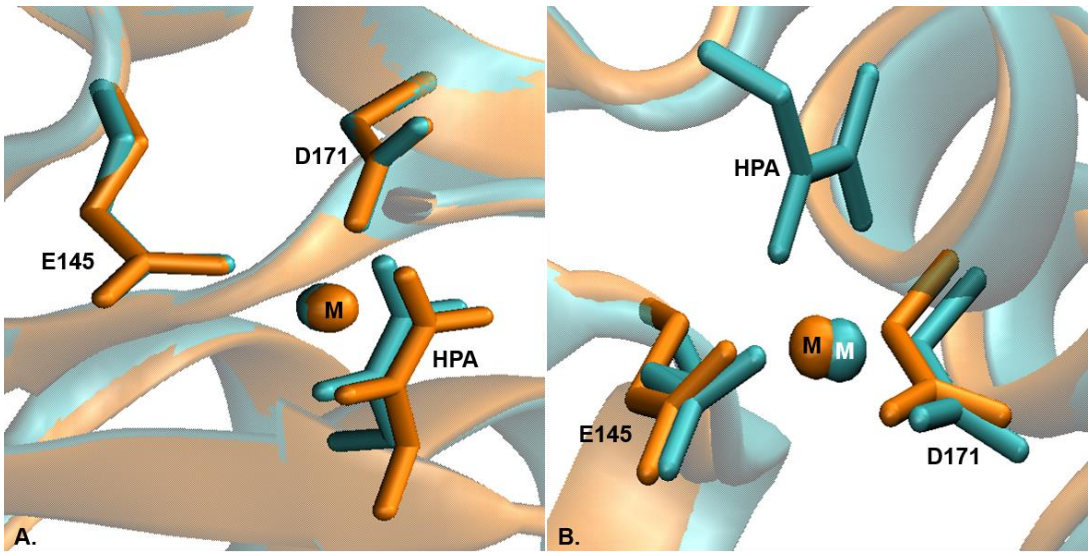


Figure 29 - *Superimposed representation of the active sites of SwHKA-Mg-HPA (cyan) and S116A-Mg-HPA (orange). Water molecules not shown. A – chain A active site of S116A-Mg-HPA; B – chain B active site of S116A-Mg-HPA.*

In chain A, where HPA is present, the active site conformations match almost exactly regarding the catalytic residues and the magnesium positions. The biggest difference is with the HPA molecule that in the S116A structure is in a planar conformation, where all atoms lie in the same plane. By contrast, in the WT structure the HPA molecule has an angular conformation. The first step of the aldol reaction by this class of proteins is the protonation of HPA at C3 (initially in angular conformation) by a water molecule, forming a planar intermediate before the approach of the aldehyde. Therefore, the difference in HPA conformation between the two structures indicates that in the S116-Mg-HPA structure the HPA molecule has already been protonated and the intermediary has been formed. As there is no aldehyde in the soaking solutions used, the difference in HPA conformation between the structures is most likely cause by the soaking times used with the two crystals, but also reveals that the aldehyde doesn't need to be in the active site for the intermediate to be formed.

In the B chain, where the magnesium ion is bound to the catalytic residues and water molecules, the structures of S116A-Mg-HPA and *Sw*HKA-Mg-HPA don't match as well as they did in chain A, with the metal ion and the side chains of E145 and D171 slightly to the side (~0.2Å). However, this difference is not significant.

If we superimpose S116A-Mg with S116A-Mg-HPA (Figure 30), we see that all the elements of the overall fold and active sites of the two chains match almost exactly, including the oxygen atoms from PEG and HPA that coordinate the magnesium.

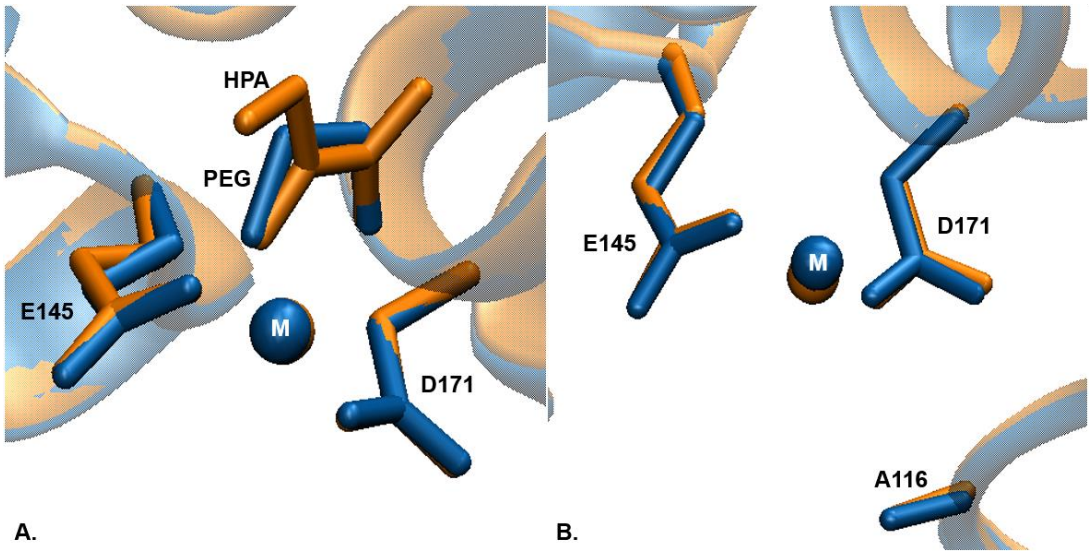


Figure 30 - *Superimposed representation of the active sites of S116A-Mg (blue) and S116A-Mg-HPA (orange). Water molecules not shown. A – chain A active site of both structures; B – chain B active site of both structures.*

In other structures of S116A-Mg not shown in this report, the coordination sphere of the magnesium ion was consistently formed by the two catalytic residues (E145 and D171) and 4 water molecules, and the occupancy of the metal was always lower than 80%. The same happened with other structures of S116A-Mg-HPA, in the active sites where HPA was not present.

The resting state was never visible in any of the *Sw*HKA-S116A structures determined until now, which leads us to conclude that the presence of S116 is essential in the formation of the resting state. Since the activity of class II pyruvate aldolases is metal dependent, having the metal in the active site at full occupancy is essential to achieve the highest possible enzymatic activity. This seems to only be possible when the resting state conformation is reached first.

3.4.2. *Bp*HKA

*Bp*HKA-Mg crystallized in the triclinic space group P1, with six molecules in the asymmetric unit. Unit cell parameters were a = 74.428Å, b = 74.558Å, c = 78.582Å, α = 101.76º, β = 92.44º and ϒ = 118.35º. The crystal used for the soaking was fished from the drop shown in Figure 19.

Good X-ray diffraction data was collected for structures of *Bp*HKA-Mg-HPA and *Bp*HKA-all but, due to the time restrictions of this project, the data has not been processed and refined yet. For this reason, these structures will not be further included in this project. Also due to the time limit of this project, the refinement of the *Bp*HKA-Mg structure has not been completed and

further refinement cycles and model building are necessary for accurate structure determination. Data collection statistics of the *Bp*HKA-Mg structure determined by X-ray crystallography in this project is shown in Table 11.

Table 11 - *Data collection statistics of BpHKA-Mg. The refinement process of this structure is not complete so no refinement statistics are shown. Parameters in parenthesis "()" correspond to outer shell values.*

	***Bp*HKA-Mg**
Space group	P 1
Resolution (Å)	75.92 - 1.70 (1.73 - 1.70)
Data completeness (%)	92.9 (93.9)
Multiplicity	1.8 (1.8)
$<I/\sigma(I)>$	8.3 (0.9)
$CC_{1/2}$	0.0.998 (0.388)
R-merge	0.048 (0.828)
R_{meas}	0.068 (1.171)
R_{pim}	0.048 (0.828)
Total reflections	262524
Unique reflections	146832

Although the refinement process of the *Bp*HKA-Mg structure is not complete at this stage, we can still draw some preliminary conclusions from the structure (Figure 31).

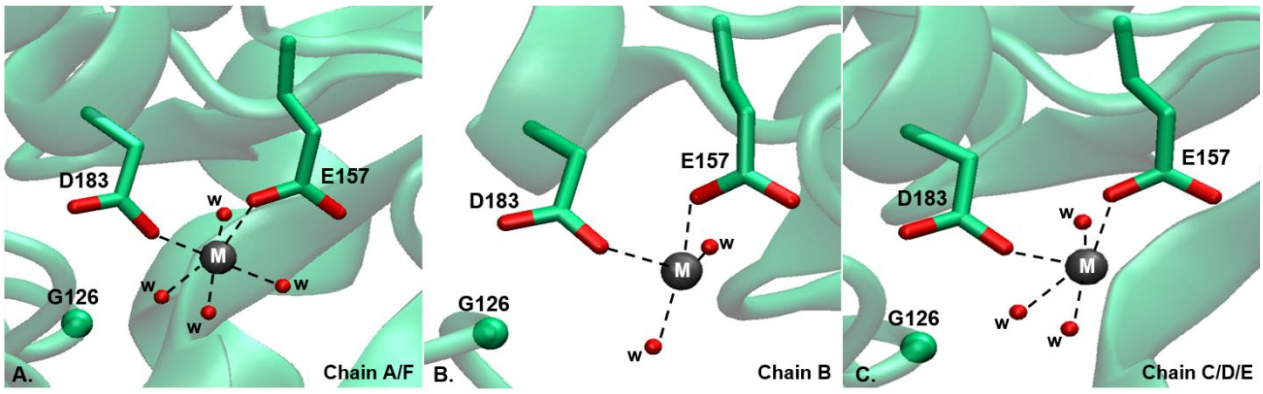


Figure 31 - *Representation of the active sites of BpHKA-Mg.*

As stated before, *Bp*HKA-Mg crystallized with six molecules in the asymmetric unit, representative of the characteristic homo-hexamer formed by class II pyruvate aldolases. More so, the six molecules formed six slightly different active sites, since in this structure the active hexametric conformation was not generated by crystalline symmetry with the monomer (*Sw*HKA-*WT*) or dimer (*Sw*HKA-S116A). Even though there are six distinct active sites, all of them are formed by 2 catalytic amino acid residues, Glu157 and Asp183, and varying numbers of water molecules (depending on the protein chain).

Represented in Figure 31 is also Gly126, the equivalent amino acid residue to Ser116 of *Sw*HKA. In the equivalent *Sw*HKA structure (*Sw*HKA-Mg), the metal ion was present at full occupancy, bound to the serine side chain. However, unlike serine, glycine does not have a side chain that could reach the active site and stabilize the magnesium ion. As a result, the coordination sphere of magnesium is formed only by the catalytic amino acids and water molecules. The mobile nature of the water molecules in the crystalline arrangement influences the presence/occupancy of the magnesium ion in the active site, since the 4 water molecules necessary are not always at the correct bond distances to coordinate and stabilize the magnesium. In *Bp*HKA-Mg's preliminary structure, this translates into magnesium occupancies lower than 1 in all active sites. In this regard, the magnesium coordination sphere and behavior in this *Bp*HKA-Mg structure is more similar to the S116A-Mg structure than the *Sw*HKA-Mg structure.

3.4.3. Comparison between SwHKA and BpHKA

To better understand the active site differences between *Sw*HKA and *Bp*HKA, the equivalent magnesium-bound structures of *Sw*HKA-Mg, S116A-Mg and *Bp*HKA-Mg were superimposed using the *Coot* function for secondary structure matching (SSM) method (Figure 32).[41]

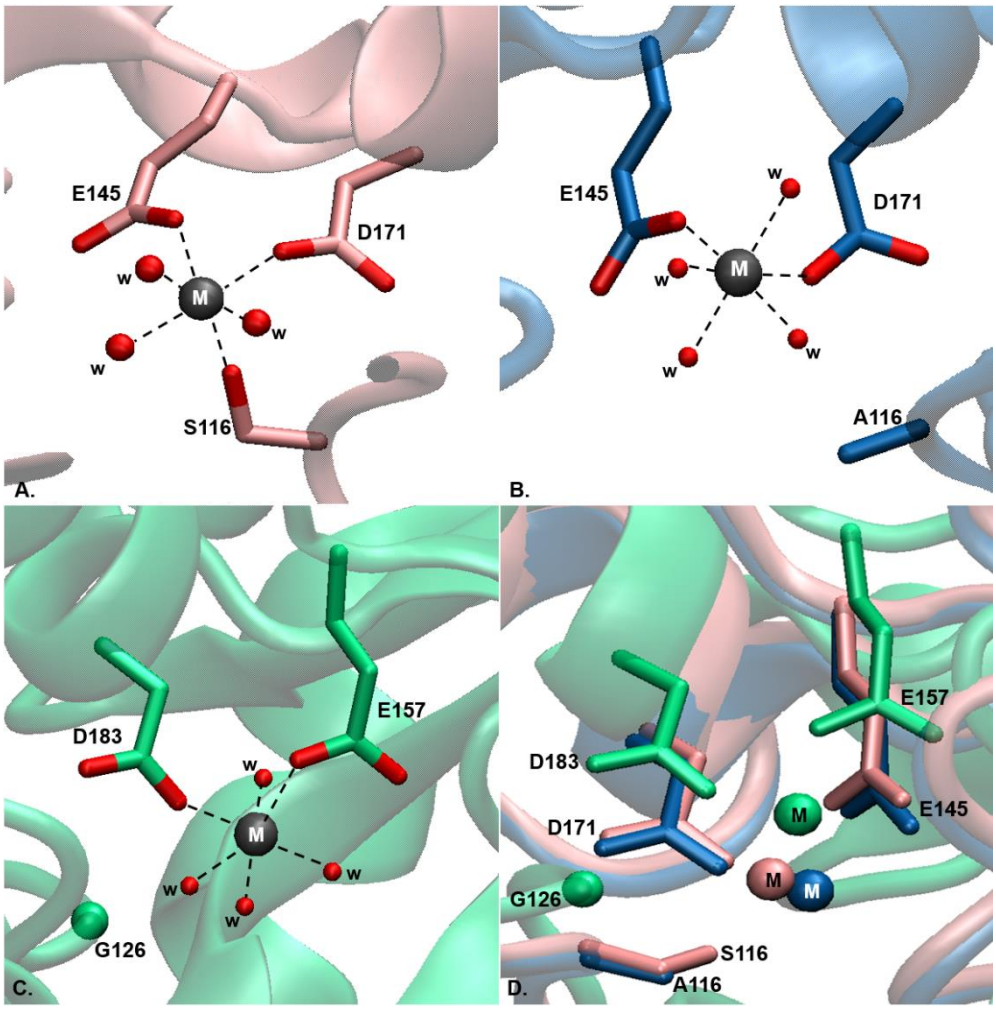


Figure 32 - *Superimposed representation of the equivalent metal (M) coordination active sites of BpHKA-Mg (green), SwHKA-Mg (pink) and S116A-Mg (blue). Water molecules (w) not shown in panel D for schematic clarity. A – SwHKA-Mg; B – Chain B of S116A-Mg; C – Chain A of BpHKA-Mg; D – BpHKA-Mg, SwHKA-Mg and S116A-Mg.*

In Figure 32 we see that the position of the catalytic amino acid residues, glutamate and aspartate, does not coincide between the *Sw*HKA and the *Bp*HKA structures. In fact, when looking at the active sites of both proteins, including the residues not represented in Figure 32, we notice some differences. Besides some punctual residue changes, the biggest difference is the fact that the active site of *Bp*HKA has a higher number of carboxylate groups turned to the active site, in the direction of the metal ion. These groups belong to the side chains of glutamate and aspartate residues positioned near the active site. The most notorious is Glu47, whose side chain is at bond distance of some water molecules of the magnesium coordination sphere. In *Sw*HKA-WT the equivalent amino acid residue, position-wise, is Gln43, whose side chain is not close enough to reach *Sw*HKA's metal center.

In both the S116A-Mg and *Bp*HKA-Mg structures, the coordination to the glutamate and aspartate residues does not seem to be enough to trap the metal ion in the active site. In these cases, the coordination sphere is completed by water molecules that due to their mobility within the solvent channels do not efficiently stabilize the metal ion either. This results in metal occupancies lower than one in both S116A-Mg and *Bp*HKA-Mg structures. *Sw*HKA-Mg however, has a third amino acid residue bound to the magnesium ion, Ser116. This serine residue not only possesses a side chain capable of coordinating to the metal ion, through the hydroxyl group, but is also in the right position in the active site to be able to reach the metal ion and bind to it in a way that forms the characteristic octahedral geometry of the magnesium coordination sphere.

As *Bp*HKA does not have a serine residue near the active site and a glycine residue is observed instead, the position of the previously mentioned Glu47 residue in *Bp*HKA's active site could be the enzyme's way of reducing the mobility of some of the water molecules that participate in the metal coordination sphere, since water molecules participating in 2 bonds tend to be less mobile than water molecules participating in only 1 bond. This residue seems to ensure the permanence of at least some of the water molecules in the positions necessary for the metal coordination sphere, which might support the fact that although some were not complete and at low occupancies, no *Bp*HKA active site was without magnesium ion or water molecules.

Simultaneously, the CCP4 supported program *GESAMT* was used for the determination of RMSD values between the 3 structures (Table 12).[46]

Table 12 - *RMSD and sequence identity values between the protein chains with equivalent metal (M) coordination active sites of SwHKA-Mg (chain A), S116A-Mg (chain B) and BpHKA-Mg (chain A).*

	***Sw*HKA-Mg *Bp*HKA-Mg**	***Sw*HKA-Mg S116A-Mg**	***Bp*HKA-Mg S116A-Mg**
RMSD	1.65	0.3434	1.66
Aligned residues	241	250	240
Seq. identity (%)	31.54	99.60	31.67

Root-mean-square deviation (RMSD) of atomic positions is a commonly used parameter of measurement of the average distance between the atoms of superimposed proteins, more specifically, the distance between the aligned Cα. As a rule, RMSD values lower than 2.0Å represent well aligned structures, with similar folds and little atomic position deviation.

As we can see in Table 12, when *Sw*HKA-Mg is aligned against *Bp*HKA-Mg, we obtain a RMSD value of 1,65Å. Although *Sw*HKA and *Bp*HKA have difference sequence lengths (251aa and 260aa, respectively) and low sequence identity between each other, a RMSD value lower than 2.0Å is not surprising since *Sw*HKA and *Bp*HKA belong to the same protein family, class II pyruvate aldolases, where the fold and overall tridimensional structure is highly conserved and no notorious conformational change upon metal or substrate binding is observed. Further structural analysis of both proteins reveals that *Bp*HKA's longer sequence mostly comes from amino acid insertions in disordered loop regions and at the c-terminal, both of which do not affect the overall protein fold.

The low RMSD value between *Sw*HKA-Mg and S116A-Mg was also to be expected, since these proteins' sequence only differs in 1 amino acid residue from each other, and no major conformational change would be expected as a result of this point mutation.

Enzymatic activity studies were also performed for *Sw*HKA-WT, SwHKA-S116A and *Bp*HKA-WT by our TU Delft collaborators and these assays showed that of the 3, *Sw*HKA-WT had the highest enzymatic activity in the presence of magnesium and other divalent metal cations. Since the only difference between *Sw*HKA-WT and *Sw*HKA-S116A is the point mutation of the Ser116 residue to an alanine residue, this seems to indicate that the WT serine residue, and the consequent metal coordination sphere formed, is essential to achieve maximum enzymatic activity. The lower enzymatic activity of *Bp*HKA-WT*,* when compared to *Sw*HKA-*WT,* could also be justified by the lack of a serine residue, or an equivalent side chain residue, in *Bp*HKA's active site.

Our working hypothesis is that Ser116 and the consequent formation of the resting state conformation is an evolutionary adaptation of *Sphingomonas wittichii* to improve its enzymatic activity when compared to other class II pyruvate aldolases: "trapping" the metal ion in the active site ensures that the protein is always ready to start the reaction when the different substrates enter the active site. Without the metal ion, even if the substrates reach the active site, no reaction occurs. With the metal ion at less-than-optimal occupancies (<1.0), the protein is active but the enzymatic activity does not reach its turnover potential.

A very superficial sequence alignment study using 53 different protein sequences found on UniProt (www.uniprot.org) with sequence identity to *Sw*HKA-*WT* and *Bp*HKA-*WT* higher than 31%, showed that the serine residue is not conserved across the sequence spectrum tested and that *Bp*HKA's equivalent amino acid residue, glycine, was more frequent in the aligned

sequences. However, a 2-dehydro-3-deoxyglucarate aldolase from *Paraburkholderia sp*. BL10I2N1, the same genus as *Bp*HKA, was identified as having a serine instead of a glycine residue. This protein, with sequence code A0A4R6F4A9, was only integrated into the UniProtKB/TrEMBL database in 2019 and as of the conclusion of this project has no known structure. The structural study of A0A4R6F4A9, and further comparison to *Sw*HKA and *Bp*HKA, could shine a light on these proteins' mechanism and metal binding behavior. Since A0A4R6F4A9 also belongs to the betaproteobacteria genus *Paraburkholderia,* a phylogenetic analysis of all known class II pyruvate aldolases could provide insight into the speciation events that lead to the differentiation between the serine and the glycine residues in the active site of these enzymes.

4. Conclusion

This project allowed first-hand contact with several techniques, used for different end-goals. SDS-PAGE gels were run to access protein sample quality and degradation degree. Differential scanning fluorimetry (nDSF) assays were used to discern protein stability, determine optimal crystal soaking concentrations and test protein thermal behavior in the presence of 3 potential aldehyde substrates. Macromolecular X-ray crystallography was used with the goal of identifying the key residues involved in the complementary stereospecificity of *Sw*HKA and *Bp*HKA and of determining the existence of a resting state conformation created upon binding of the metal ion to a different coordination sphere than the one observed in the active state of the enzyme.

The SDS-PAGE gels revealed stable and pure *Sw*HKA-WT and *Sw*HKA-S116A protein samples and a degraded *Sw*HKA-S116C protein sample. This technique was also used for long-term incubation experience with *Bp*HKA, and revealed that this protein degraded overtime at 4ºC, which unknowingly affected some crystallization trials.

In the nDSF assays performed, *Sw*HKA and *Bp*HKA showed different thermal behaviors after incubation with $MgCl_2$ and hydroxypyruvate (HPA). While *Sw*HKA reached maximum T_m values in the presence of magnesium alone, with a total temperature increase of 13.5ºC, *Bp*HKA's highest T_m shift (~25ºC) was reached in the presence of both magnesium and HPA, as well as very high concentrations of magnesium. The thermal stability of both proteins was also tested in the presence of 3 potential aldehyde substrates: DL-glyceraldehyde, glycolaldehyde and butyraldehyde. In this assay, *Sw*HKA showed the same thermal behavior towards all aldehydes tested, on the other hand *Bp*HKA showed a clear preference for DL-glyceraldehyde. The use of the nDSF methodology also allowed us to the determine the initial soaking concentrations of $MgCl_2$, HPA and DL-glyceraldehyde as 100mM of each, although many more variations were used.

Although the crystallization conditions of *Sw*HKA were already known at the beginning of this project, crystallization conditions for *Sw*HKA mutants S116A and S116C, as well as *Bp*HKA-WT, were unknown. Therefore, the crystallization conditions of these proteins were determined making use of commercially available crystallizations screens and further extensive manual optimization and fine-tuning of the hit conditions. After good X-ray diffracting crystals of each protein had been obtained, X-ray data was collected. After data processing, structure solution and several cycles of refinement and model building, structural determination was achieved for

*Sw*HKA-WT (*Sw*HKA-apo and *Sw*HKA-all), *Sw*HKA-S116A (S116A-Mg and S116A-Mg-HPA) and *Bp*HKA-WT (*Bp*HKA-Mg, structure refinement incomplete due to limited time).

The resting state was never isolated in any of the *Sw*HKA-S116A structures determined, which leads us to conclude that the presence of S116 is indispensable to the formation of the resting state. Since the activity of class II pyruvate aldolases is metal dependent, having the metal in the active site at full occupancy is essential to achieve the highest possible enzymatic activity. For *Sw*HKA, this is only achieved when the resting state conformation is reached first. Preliminary *Bp*HKA structural data also showed that this protein does not form the resting state conformation in the presence of magnesium due to the lack of a serine residue in the appropriate place. As a result, *Bp*HKA had lower metal occupancy and enzymatic activity than *Sw*HKA, which does form the resting state conformation.

In conclusion, we were unsuccessful in determining the cause of the stereo-complementarity between *Sw*HKA and *Bp*HKA, but successfully confirmed the importance of a serine residue (S116) in the active site of these proteins. If *Sw*HKA is ever applied to the pharmaceutical industry as a pathway to the production of ketoacids, the S116 residue should be preserved to guarantee the best enzymatic activity.

In the future, it would be interesting to study a *Bp*HKA mutant where the WT glycine is mutated to a serine residue, mimicking the active site composition of *Sw*HKA, and see how the mutation affects active site occupancy of the magnesium cation and overall enzymatic activity. Time-resolved experiments and the structural study of *Sw*HKA and *Bp*HKA in the presence of other aldehydes could also provide a better understanding of the stereoselective mechanism of these enzymes.

5. References

1. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.* **280**, 5705–5736 (2013).

2. Wlodawer, A., Dauter, Z. & Jaskolski, M. *Protein Crystallography: Methods and Protocols*. (Springer Nature, 2017).

3. Mcpherson, A. & Gavira, J. A. Introduction to protein crystallization. 2–20 (2014). doi:10.1107/S2053230X13033141

4. McPherson, A. & Cudney, B. Optimization of crystallization conditions for biological macromolecules. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **70**, 1445–1467 (2014).

5. Chayen, N. E. & Saridakis, E. Protein crystallization: From purified protein to diffraction-quality crystal. *Nat. Methods* **5**, 147–153 (2008).

6. Chayen, N. E. Turning protein crystallisation from an art into a science. *Curr. Opin. Struct. Biol.* **14**, 577–583 (2004).

7. Hassell, A. M. *et al.* Crystallization of protein – ligand complexes. 72–79 (2007). doi:10.1107/S0907444906047020

8. Muller, I. Guidelines for the successful generation of protein – ligand complex crystals. **7983**, 79–92 (2017).

9. Bergfors, T. Seeds to crystals. *J. Struct. Biol.* **142**, 66–76 (2003).

10. Pflugrath, J. W. Practical macromolecular cryocrystallography. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **71**, 622–642 (2015).

11. Garman, E. 'Cool' crystals: Macromolecular cryocrystallography and radiation damage. *Curr. Opin. Struct. Biol.* **13**, 545–551 (2003).

12. Aggarwal, M. *et al.* Effects of cryoprotectants on the structure and thermostability of the human carbonic anhydrase II-acetazolamide complex. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 860–865 (2013).

13. Deutsches Elektronen-Synchrotron - Photon Science. How does a synchrotron radiation source work? (2019). Available at: http://photon-

science.desy.de/research/students__teaching/primers/synchrotron_radiation/index_eng.html. (Accessed: 12th April 2021)

14. Weiss, M. S. Global indicators of X-ray data quality. *J. Appl. Crystallogr.* **34**, 130–135 (2001).

15. Blow, D. *Outline of Crystallography for Biologists*. (Oxford University Press, 2002).

16. Taylor, G. L. Introduction to phasing. *Acta Crystallogr. Sect. D* **66**, 325–338 (2010).

17. Shabalin, I. G., Porebski, P. J. & Minor, W. Refining the macromolecular model–achieving the best agreement with the data from X-ray diffraction experiment. *Crystallogr. Rev.* **24**, 236–262 (2018).

18. Clapés, P., Fessner, W., Sprenger, G. A. & Samland, A. K. Recent progress in stereoselective synthesis with aldolases. *Curr. Opin. Chem. Biol.* **14**, 154–167 (2010).

19. Baker, P. & Seah, S. Y. K. Rational design of stereoselectivity in the class II pyruvate aldolase BphI. *J. Am. Chem. Soc.* **134**, 507–513 (2012).

20. Coincon, M., Wang, W., Sygusch, J. & Seah, S. Y. K. Crystal Structure of Reaction Intermediates in Pyruvate: Substrate Cleavage, Enolate Stabilization and Substrate Specificity. *J. Biol. Chem.* **287**, 36208–36221 (2012).

21. Walters, M. J. *et al.* Characterization and Crystal Structure of Escherichia coli KDPGal Aldolase. *Bioorg. Med. Chem. Lett.* **16**, 710–720 (2008).

22. Drueckhammer, D. G. *et al.* EnzymeCatalysis in Synthetic Carbohydrate Chemistry. *Synthesis (Stuttg).* (1991).

23. Blackwell, N. C., Cullis, P. M., Cooper, R. A. & Izard, T. Rhombohedral crystals of 2-dehydro-3- deoxygalactarate aldolase from Escherichia coli. *Acta Crystallogr. Sect. D* **55**, 1368–1369 (1999).

24. Schurmann, M. & Sprenger, G. A. Fructose-6-phosphate Aldolase Is a Novel Class I Aldolase from Escherichia coli and Is Related to a Novel Group of Bacterial Transaldolases. *J. Biol. Chem.* **276**, 11055–11061 (2001).

25. Marsh, J. J. & Lebherz, H. G. Fructose-bisphosphate aldolases: an evolutionary history. *Trends Biochem. Sci.* 110–113 (1992).

26. Rea, D., Fülöp, V., Bugg, T. D. H. & Roper, D. I. Structure and Mechanism of HpcH: A

Metal Ion Dependent Class II Aldolase from the Homoprotocatechuate Degradation Pathway of Escherichia coli. *J. Mol. Biol.* **373**, 866–876 (2007).

27. De Berardinis, V. *et al.* Expanding the reaction space of aldolases using hydroxypyruvate as a nucleophilic substrate. *Green Chem.* **19**, 519–526 (2017).

28. Laurent, V. *et al.* Pyruvate Aldolases Catalyze Cross-Aldol Reactions between Ketones: Highly Selective Access to Multi-Functionalized Tertiary Alcohols. *ACS Catal.* **10**, 2538–2543 (2020).

29. Marsden, S. R. *et al.* CH-π Interactions Promote the Conversion of Hydroxypyruvate in a Class II Pyruvate Aldolase. *Adv. Synth. Catal.* **361**, 2649–2658 (2019).

30. Marsden, S. R. *et al.* Thermodynamics Determine the Diastereochemical Outcome of Catalytic Reactions. *ChemCatChem* 2530–2536 (2021). doi:10.1002/cctc.202100178

31. Lehninger, N. D. & COX., M. M. *Lehninger Principles of Biochemistry*. (Worth Publishers, 2000).

32. Svasti, J. & Panijpan, B. SDS-polyacrylamide gel electrophoresis (SDS-PAGE). *Curr. Protoc. Essent. Lab. Tech.* **54**, 560–562 (1977).

33. Strutz, W. Exploring Protein Stability by NanoDSF. *Biophys. J.* **110**, 393a (2016).

34. NanoTemper Technologies. Prometheus NT.48 - Product Information. (2015).

35. Cianci, M. *et al.* P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. *J. Synchrotron Radiat.* **24**, 323–332 (2017).

36. Kabsch, W. research papers XDS research papers. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).

37. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D* 1204–1214 (2013). doi:10.1107/S0907444913000061

38. Potterton, L. *et al.* CCP 4i2 : the new graphical user interface to the CCP 4 program suite. *Acta Crystallogr. Sect. D* 68–84 (2018). doi:10.1107/S2059798317016035

39. Vagin, A. & Teplyakov, A. MOLREP: An Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).

40. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255 (1997).

41. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

42. Uski, V., Krissinel, E. & Ballard, C. CCP4 Web Services and Cloud Computing Developments. *Acta Crystallogr. Sect. A* (2017).

43. Miotto, M. *et al.* Insights on protein thermal stability: A graph representation of molecular interactions. *Bioinformatics* **35**, 2569–2577 (2019).

44. Pucci, F., Bourgeas, R. & Rooman, M. Predicting protein thermal stability changes upon point mutations using statistical potentials : Introducing HoTMuSiC. *Nat. Publ. Gr.* **6**, 1–9 (2016).

45. Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E. & Bell, J. A. Extent and Nature of Contacts Between Protein Molecules in Crystal Lattices and Between Subunits of Protein Oligomers. **514**, 494–514 (1997).

46. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).

6. Supplementary data

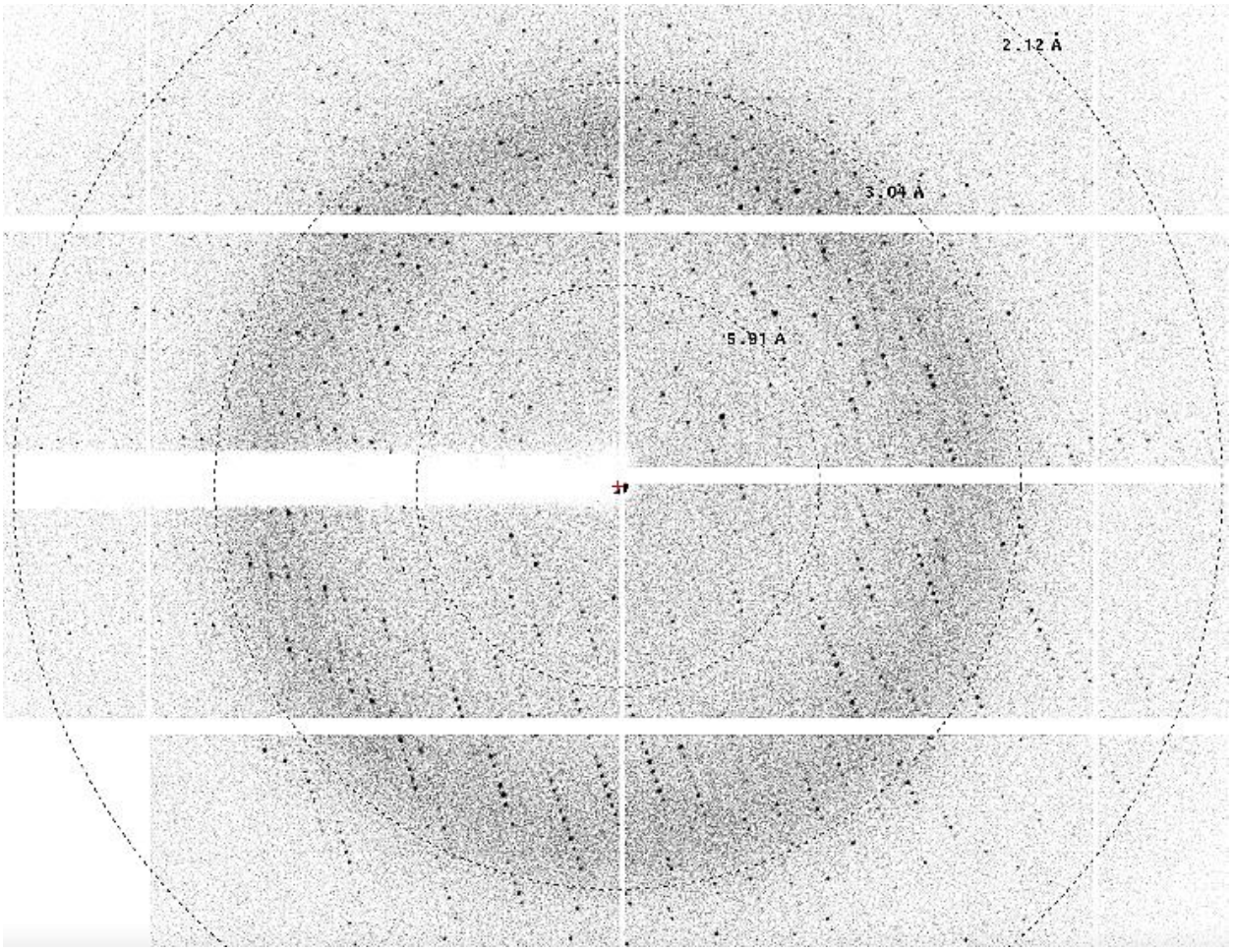
S1 - Stock reagents and equipment used. Solutions used in the report but not mentioned in this table were pre-done by the SPC staff.

	Name	Purity (%)	Manufacturer
Reagents	Magnesium Chloride	≥99.0	Sigma-Aldrich
	Glycerol	≥99.5	Carl Roth®
	Potassium phosphate monobasic	≥99.0	Sigma-Aldrich
	Potassium phosphate dibasic	≥99.0	Sigma-Aldrich
	Sodium phosphate monobasic monohydrate	≥99.5	Sigma-Aldrich
	Sodium phosphate dibasic	≥99.0	Sigma-Aldrich
	HEPES	≥99.5	Carl Roth®
	Bis-Tris	≥99.0	Carl Roth®
	Bis-Tris propane	≥99.0	Carl Roth®
	Pyruvic acid	≥98.0	Sigma-Aldrich
	Sodium β-hydroxypyruvate hydrate	≥97.0	Sigma-Aldrich
	DL-glyceraldehyde	≥90.0	Sigma-Aldrich
	Glycolaldehyde dimer	-	Sigma-Aldrich
	Butyraldehyde	≥96.0	Sigma-Aldrich
Equipment	**Name**	**Manufacturer**	
	Prometheus NT.48 nanoDSF	Nanotemper	
	Scorpion Screen Builder	ARI	
	NanoDrop 2000	ThermoScientific	
	Ultracentrifuge Optima XPN-80	Beckman	
	Mosquito-LCP	TTPLabtech	
	Liquidador96	Mettler Toledo	
	The RockImager-1000	Formulatrix	
	Swissci plates	TTPLabtech	
	CrystalDirect plates	MiTeGen	
	SevenEasy Standart pH Meter Line	Mettler Toledo	
	Leica MZ16 A stereomicroscope	Leica Microsystems	

S2 - Overview of magnesium dissociation constants (K_d) for *Sw*HKA (WT and mutants) and *Bp*HKA.

Protein		K_d (μM)
***Sw*HKA**	WT	58.4±4.1
	S116A	37.9±3.0
	S116C	80.4±19.4
BpHKA - WT		11.7 ± 0.5

S3 - High quality X-ray diffraction pattern.



S4 - Low quality X-ray diffraction pattern.

