BÁRBARA BEATRIZ DA COSTA BOTELHO SOUSA

Bachelor degree in Biology

CRYSTAL STRUCTURE OF BONE MARROW TYROSINE KINASE FROM CHROMOSOME X WITH COVALENT LIGANDS

Dissertation for the Master degree in Biochemistry for Health

Supervisor: Dr. Pedro M. Matias, ITQB-NOVA

Co-supervisor: Dr. Marta C. Marques, iMM

September, 2018

BÁRBARA BEATRIZ DA COSTA BOTELHO SOUSA

Bachelor degree in Biology

CRYSTAL STRUCTURE OF BONE MARROW TYROSINE KINASE FROM CHROMOSOME X WITH COVALENT LIGANDS

Dissertation for the Master degree in Biochemistry for Health

Supervisor: Dr. Pedro M. Matias, ITQB-NOVA

Co-supervisor: Dr. Marta C. Marques, iMM

Júri:

Presidente: Prof. Doutora Teresa Catarino

Arguente: Prof. Doutora Teresa Santos-Silva

Vogal: Prof. Doutora Marta Coimbra Marques

Instituto de Tecnologia Química e Biológica

September, 2018

TABLE OF CONTENTS

Abstract i

Resumo ii

List of figures iii

List of tables vii

List of abbreviations viii

Chapter 1 9

1. Introduction 9

1.1. Prostate cancer 9

1.1.1. Epidemiology 9

1.1.2. Pathology, genetics and risk factors 9

1.2. Introducing Bone Marrow Tyrosine Kinase in chromosome X 10

1.2.1. Three-dimensional structure of BMX catalytic domain 10

1.2.2. BMX expression and its role in prostate cancer 11

1.3. BMX kinase inhibitors 12

1.4. BMX production using the baculovirus-insect cell expression system 12

1.4.1. Baculovirus biology and life cycle: *in vivo* vs. *in vitro* cultures 13

1.4.2. Baculovirus-insect cell expression system 14

1.4.3. Generating the recombinant bacmid in *E. coli* cells 15

1.4.4. Bioprocess engineering aspects of insect cell culture 16

1.5. Sample quality assessment 17

1.5.1. Dynamic light scattering (DLS) 17

1.5.2. Circular Dichroism (CD) 17

1.6. Protein-ligand complex characterization 18

1.6.1. Differential scanning fluorimetry (DSF) 18

1.6.2. Surface Plasmon Resonance (SPR) 19

1.7. X-ray crystallography for protein-ligand structure determination 21

1.7.1. Protein crystallization 21

1.7.2. X-ray sources and data collection 23

1.7.3. Protein structure determination 25

1.7.4. Refinement and cross-validation 26
1.8. Scope of the thesis 27

Chapter 2 29

2. Methodology 29
2.1. Synthesis of the recombinant donor plasmid 29
2.2. Transformation of competent *E. coli* cells 29
2.2.1. Analysis of the recombinant bacmid DNA by PCR 30
2.3. Baculovirus-insect cell expression vector system 30
2.3.1. Insect cell culture 30
2.3.2. Insect cell transfection 30
2.3.3. Amplification of the viral stock 31
2.3.4. Virus titration by CGCA and MTT methods 31
2.4. hBMX production 32
2.4.1. Small scale feasibility test and quality assessment 32
2.4.2. Large-scale production of hBMX in stirred-tank bioreactors 32
2.4.3. Downstream processing 33
2.4.4. Purified sample quality assessment by DLS and CD 34
2.5. Protein-ligand complex characterization 35
2.5.1. Differential scanning fluorimetry (DSF) 35
2.5.2. Surface Plasmon resonance (SPR) 35
2.6. Protein-ligand complex determination by X-ray crystallography 36
2.6.1. hBMX co-crystallization with a covalent binder 36
2.6.2. Model building and refinement 36

Chapter 3 39

3. Results and discussion 39
3.1. Analysis of the transformed competent *E. coli* cells by PCR 39
3.2. Insect cell transfection and protein production 40
3.3. Purification process 44
3.3.1. Purified sample quality assessment by DLS and CD 47

3.4. Protein-ligand complex characterization by DSF and SPR 48
3.5. Protein-ligand complex structure determination by X-ray crystallography 51
3.5.1. hBMX crystallization 51
3.5.2. Overall structure of hBMX 52
3.5.3. Interactions of BMX-CI-651 with hBMX catalytic domain 55

Chapter 4 59
4. Conclusions and future perspectives 59
Acknowledgments 61
Bibliography 62

Abstract

Bone marrow tyrosine kinase in chromosome X (BMX) is a major member of the TEC family kinases and has been implicated in tumorigenecity, motility, proliferation and differentiation. BMX is highly overexpressed in prostate cancer and it is involved in the adaptive compensatory mechanism of castrate-resistance prostate cancer to androgen deprivation therapy. Besides, it suppresses a core component of the intrinsic apoptotic pathway, granting tumor cells the ability to escape apoptosis induced by chemotherapeutic drugs. BMX knockout mouse have a normal lifespan, without an obvious altered phenotype, suggesting that therapies based on BMX inhibition, may have limited side effects. We developed a series of BMX-IN-1 analogues that showed an increased inhibitory capacity when compared to BMX-IN-1. Since crystallographic information is essential to understand the molecular basis of BMX inhibition by covalent inhibitors we establish a baculovirus-insect expression system protocol for the production of the recombinant human BMX. Here we report the full biochemical and biophysical characterization of human BMX alone and in complex with different covalent ligands. Evaluation of the protein sample quality was performed prior to crystallization trials, using SDS-PAGE, Western Blot, Dynamic Light Scattering, and Circular Dichroism, and BMX-ligand complex characterization was carried out by Differential Scanning Fluorimetry (DSF) and Surface Plasmon Resonance (SPR). Moreover, we set up crystallization trials to generate crystals big enough for X-ray diffraction analysis. From DSF and SPR studies we identified 3 hit compounds capable of stabilizing the hBMX over 9ºC, and with greater affinities towards BMX than BMX-IN-1. Crystallographic information of the hBMX/BMX-CI-651 complex revealed that target specificity is not only achieved via modulation of the Michael acceptor moiety, but also via other substituents (e.g., in the pyridinone ring) which contribute to non-covalent interactions. Hopefully, these findings will play a key role in studies aimed at developing improved therapeutic molecules and to help overcome the current limitations in prostate cancer therapy.

Keywords: human recombinant BMX, baculovirus-insect expression vector system, covalent binders, biochemical and biophysical assays, X-ray crystallography structure determination

Resumo

A bone marrow tyrosine kinase in the X chromossome (BMX) é uma proteína pertencente à família das TEC kinases e encontra-se implicada na tumorigénese, motilidade, proliferação e diferenciação tumoral. BMX é sobreexpressa no cancro da próstata e está envolvida no mecanismo compensatório adaptativo do cancro a terapias de privação androgénica. Além disso, a sua sobreexpressão suprime um componente central da via apoptótica intrínseca, conferindo às células tumorais a capacidade de sobreviver à apoptose induzida por drogas quimioterapêuticas. Ratos homozigóticos com expressão de BMX suprimida apresentam uma expectativa de vida normal, sugerindo que terapias baseadas na inibição de BMX poderão ter efeitos secundários limitados. Com base nesta informação, desenvolvemos um conjunto de análogos do composto BMX-IN-1 que em comparação mostraram uma maior capacidade inibitória. Uma vez que a informação cristalográfica é essencial para compreender o mechanismo de inibição da BMX, desenvolvemos um sistema de expressão de inseto-baculovírus para a produção desta proteína. Aqui relatamos a caracterização bioquímica e biofísica completa da própria BMX recombinante humana e em complexo com diferentes ligandos. Avaliámos a qualidade da amostra purificada previamente à cristalização, utilizando SDS-PAGE, Western Blot, Dynamic Light Scattering e Circular Dichroism, e a caracterização do complexo BMX-ligando foi realizada por Differential Scanning Fluorimetry (DSF) e Surface Plasmon Resonance (SPR). Efectuámos também ensaios de cristalização de modo a gerar cristais para a análise de difração de raios-X. Por DSF e SPR, identificámos 3 compostos capazes de estabilizar a hBMX acima de 9ºC, e com maior afinidade para a hBMX do que BMX-IN-1. Determinação estrutural do complexo hBMX/BMX-CI-651 revelou que a especificidade ao alvo não é apenas alcançada via modulação do aceitador de Michael, mas também via outros substituintes (e.g., no anel de piridinona) que contribuem como ligações não covalentes. Este conhecimento servirá como base para estudos que visam o desenvolvimento de moléculas terapêuticas, de modo a superar as limitações existentes nas terapias atuais.

Palavras-chave: BMX proteína recombinante humana, sistema de vetores de expressão de baculovírus-células de insecto, ligantes covalentes, ensaios bioquímicos e biofísicos, determinação de estrutura por cristalografia de raios-X

List of figures

Figure 1.1. Sequence similarity of the TEC family members. PH, pleckstrin homology; CS, cystein string; BH, BTK homology; PR, proline rich; TH, TEC homology; SH, Src homology (Adapted from ref. 12). 10

Figure 1.2. Structure of human BMX catalytic domain. A) The overall catalytic domain of hBMX tyrosine kinase with the amino- lobe colored in teal blue and the carboxy- lobe colored in grey. B) Important structural segments highlighted: the G-loop (marine blue); the hinge region (cyan blue); the αC-helix (green); the DFG motif (yellow); and the activation loop (red) (Adapted from ref. 15). 11

Figure 1.3. Chemical structure of BMX-IN-1.[28] 12

Figure 1.4. The replication cycle of a baculovirus.[30] 13

Figure 1.5. Generation of recombinant baculovirus bacmid and gene expression with the Bac-to-Bac® Expression System.[43] 15

Figure 1.6. Origin of the circular dichroism effect and protein CD spectra. (A) The left (L) and (R) components of plane polarized light: (I) the two components have the same amplitude and the combination regenerates a plane polarized plane; (II) the two components have different magnitudes and the combination of the two results in an elliptically polarized light. (B) Far UV spectra for different types of secondary elements of proteins. Solid line, α-helix; long dashed lined, anti-parallel β-sheet; dotted line, type I β-turn; cross dashed line, extended 3_1-helix or poly (Pro) II helix; short dashed line, irregular structure (Adapted from ref. 52). 18

Figure 1.7. Representation of a protein melting curve in the presence of a fluorescent dye. The fluorescence intensity is plotted as a function of the temperature for the unfolding protein.[59] ... 19

Figure 1.8. Principle of a SPR detection experiment.[62] 20

Figure 1.9. A typical binding cycle observed with an SPR sensorgram. The dotted lines represent the running buffer, while the solid line indicates the injection of an analyte, resulting in the binding of the analyte to a receptor molecule immobilized on the chip surface. [63] 20

Figure 1.10. Representation of one unit cell. The position of an atom in the unit cell determines a specific set of spatial coordinates x, y, z. [64] 21

Figure 1.11. Representation of a protein crystallization phase diagram indicating different methodologies to reach the nucleation and metastable zones.[70] 22

Figure 1.12. Diagram of an X-ray diffraction experiment. An X-ray beam generated by an X-ray tube is directed at a protein crystal. The result is a diffraction pattern obtained from X-ray scattering that is recorded on a detector (film). The intensities and positions of each reflection contain the information required to reconstruct the electron density map.[64] 24

Figure 1.13. Schematic representation of a synchrotron (Adapted from ref. 74). 25

Figure 1.14. Schematic representation of Bragg's diffraction model. The ray planes R1 and R2 interfere constructively if the additional distance of R2 (d_{hkl}) is an integral number of wavelengths: $2d \sin \theta = n \lambda$.[64] 25

Figure 2.1. Recombinant donor plasmid map containing the hBMX gene sequence (provided by the Invitrogen GeneArt™)..... 29

Figure 2.2. Large-scale production of hBMX in a 5 L stirred-tank bioreactor..... 33

Figure 3.1. Plated DH10Bac™ E. coli cells after transformation proceedings. A) Transformation content concentrated to 1x; B) transformation content concentrated to 100x. 39

Figure 3.2. Analysis of the recombinant bacmid DNA. The DNA templates from both white (W) and blue (B) colonies used in the PCR reaction, were run on a 1% agarose gel in parallel to GeneRuler™ 1 kb Plus DNA marker (M). The PCR reaction of the recombinant bacmid DNA from the white colony resulted in amplification of a fragment with approximately 1311 bp – indicating an appropriate transposition of the hBMX gene. On the other hand, the bacmid DNA from the blue (negative control) was not amplified..... 40

Figure 3.3. Microscopic view of Sf-9 cells transfected with the recombinant bacmid DNA. Cell shape, size and density was monitored immediately right after transfection (0 hpt) and for several hours post-transfection (24, 48, and 72 hpt). A negative control consisting of non-infected cells was also performed. Note the increase in cell concentration and nuclei size at 72 hpt, as well as the difference in cell density between the transfected and non-transfected cells. 41

Figure 3.4. Graphic illustrations of the daily monitoring of some cytophatic effects caused by the baculovirus infection. A) Viable cell concentration ($x10^6$ cells/mL); B) Percentage of cell viability; C) Cell size (µm)..... 43

Figure 3.5. SDS-PAGE gel and Western Blot for hBMX detection. The SDS-PAGE and Western Blot were performed using the pellet of cells infected with a MOI of 0.01 pfu/cell (lanes 1 and 2) and of 0.001 pfu/cell (lanes 3 and 4). The presence of hBMX in cell lysate was confirmed at 48 hpi (lanes 1 and 3) and at 72 hpi (lanes 2 and 4). SDS-page was run on a 4-15% acrylamide gel in parallel with Precision Plus Protein™ All Blue Prestained Protein Standards (M). A positive control was also performed by loading a 6xHis-tag protein into the gel (C+). The Western Blot was carried out on the baculovirus infected Sf-9 cells with the 6x-His Tag mouse primary antibody and with the goat anti-Mouse IgG (H+L) secondary antibody. 43

Figure 3.6. Graphic illustrations of the daily monitoring of the cells in the bioreactor. A) Viable cell concentration (x10^6 cells/mL); B) Percentage of cell viability; C) Cell size (µm). Positive (infected) and negative (non-infected) controls were also performed. 44

Figure 3.7. Size exclusion chromatography of the hBMX protein and SDS-PAGE gel analysis. A) Size exclusion chromatography. Only a single peak was observed. B) SDS-PAGE gel analysis. Lanes 15 to 91 denote the peak fractions obtained during the chromatographic experiment. 45

Figure 3.8. Affinity chromatography of the 6xHis-hBMX protein after TEV digestion and Western Blot analysis. A) Affinity chromatography. Protein was eluted with different percentages of imidazole, resulting in 6 peaks: peak U corresponds to the untagged hBMX; the remaining peaks correspond to the 6xHis-tagged protein. B) Western Blot analysis. Lanes U to 66 denote the peak fractions obtained during the chromatographic experiment. The untagged protein is not present in lane U, whereas the 6xHis-tagged protein is observed on the lanes 8 to 66. 45

Figure 3.9. Analysis of the product obtained in the end of the purification process by Western Blot. A) Untagged hBMX protein. B) 6xHis-tagged hBMX protein. Note that the untagged protein is contaminated by the 6xHis-tagged protein and there is also some degradation in the 6xHis-tagged protein sample. 46

Figure 3.10. DLS autocorrelation function and spectrum of hBMX, where the relative intensity (%) is displayed as a function of the diameter of the protein particle. A) Analysis carried out at on a sample not ultracentrifuged. B) Analysis of the same sample after ultracentrifugation. 47

Figure 3.11. Characterization of hBMX by circular dichroism. The protein displays a curve typical of predominantly α-helical structures. 48

Figure 3.12. Thermal melting derivative curves of hBMX in the presence of 4 different covalent binders: (a) BMX-IN-1; (b) BMX-CI-651; (c) BMX-CI-51; (d) BMX-CI-61. The ΔT_m represents the shift in the melting temperature curve caused by the presence of a binder molecule. 49

Figure 3.13. SPR binding analysis of hBMX catalytic domain with 4 different covalent compounds: (a) BMX-IN-1 (reference compound); (b) BMX-CI-651; (c) BMX-CI-51; (d) BMX-CI-61. Each graphic depicts the response upon the injection (t=0 seconds) of a compound at 5 different dilutions, and the response of subsequent washing step (t=200 seconds). The raw data of individual measurements are shown in color and the fitted data in black. Note that the compounds BMX-CI-651, BMX-CI-51, and BMX-CI-61, do not dissociate after the washing step. 50

Figure 3.14. Optical microscopy image of hBMX crystals in complex with BMX-CI-651. 52

Figure 3.15. The overall structure of the hBMX kinase catalytic domain in complex with BMX-CI-651. The crystal structure of hBMX is represented as teal (for the amino-lobe) and grey (for the carboxy-lobe) cartoon. The covalent inhibitor is colored magenta and shown in ball and sticks representation. 53

Figure 3.16. Alignment of the hBMX kinase catalytic domain structures. A) Superposition of the two independent chains of the hBMX/BMX-CI-651 complex structure. Chain A is colored in marine blue and chain B in grey. B) Superposition of hBMX/BMX-CI-651 complex with BMX/PP2. Chain A of hBMX/BMX-CI-651 is colored in marine blue and the BMX/PP2 complex is represented in orange. 53

Figure 3.17. Positioning of the DFG Aspartic acid and Phenylalanine in the active and inactive kinase conformations. A) Close-up of the DFG Aspartic acid and Phenylalanine in apo BTK (PDB entry: 3P08). The DFG Asp540 points at the ATP binding pocket and the DFG Phe445 residue is positioned in the back cleft and facilitates the formation of the β3 Lys430-αC Glu455 ion pair. B) Close-up of the DFG Aspartic acid and Phenylalanine in the hBMX/BMX-CI-561 complex. The DFG Asp554 points down and away from the ATP binding pocket and the Phe555 is packed against the αC-helix and swings up, blocking the ion pair formation. The Lys445 also interacts with BMX-CI-651. 55

Figure 3.18. Binding mode of BMX-CI-651 to hBMX kinase catalytic domain. The covalent bond between the BMX-CI-651 and the nucleophilic Cys496 is showed as a yellow solid line. Other polar (non-binding) interactions of BMX-CI-651 with the residues Lys445 and Ile492 are represented as red dash lines. 56

Figure 3.19. Hydrophobic interactions of the inhibitor BMX-CI-651 within the hBMX catalytic domain. The interactions with the residues Tyr491, Ala443, Val431, and Leu543 are represented as blue dash lines. 57

List of tables

Table 2.1. Primers sequence and PCR reaction conditions. 30

Table 2.2. Data processing statistics for the crystallographic structure of hBMX in complex with BMX-CI-651. 37

Table 2.3. Refinement statistics for the crystallographic structure of hBMX in complex with BMX-CI-65. 38

Table 3.1. Baculovirus titers obtained with CGCA and MTT methods 41

Table 3.2. Summary of purification parameters of the recombinant hBMX. 46

Table 3.3. SPR binding parameters for the interaction of hBMX with different covalent binders. 51

Table 3.4. Superposition of hBMX/BMX-CI-651 with hBMX/Dasatinib and hBMX/PP2 complex structures. 54

List of abbreviations

θ – Ellipticity
λ – Wavelength
$|F_{hkl}|$ – Structure factor amplitude
ϕ_{hkl} – Phase angle
ΔT_m – Shift in the melting temperature
AcMNPV – *Autographa californica* multiple nuclear polyhedrosis virus
ADT – Androgen-deprivation therapy
Akt – Protein kinase B
APE – Alanine-proline-glutamic acid
Bacmid – Baculovirus shuffle vector
BAK – Bcl-2 homologous antagonist
BMX – Bone marrow tyrosine kinase gene from the X chromosome
BTK – Bruton's tyrosine kinase
BVs – Budded virions
CD – Circular dichroism
CGCA – Cell growth cessation assay
CRPC – Castrate-resistant prostate cancer
DFG – Aspartic acid-phenylalanine-glycine
DLS – Dynamic light scattering
DSF – Differential scanning fluorimetry
FAK – Focal adhesion kinase
G-loop – Gly-X-Gly-X-X-Gly
HA – Heavy atom
hBMX – Human recombinant BMX
hpi – Hours post-infection
hpt – Hours post-transfection
IC-BEVS – Baculovirus-insect cell expression vector system
IL-6 – Interleukin-6
ITK – Interleukin (IL)-2-inducible T-cell kinase
k_D – Affinity constant
k_{off} – Rate of dissociation
k_{on} – Rate of association
MAD – Multiple wavelength anomalous dispersion
MIR – Multiple isomorphous replacement
MOI – Multiplicity of infection
MR – Molecular replacement
MTT – Microculture tetrazolium
OBs – Occlusion bodies
ODVs – Occlusion-derived virions
pfu – Plaque forming units
PH – Pleckstrin homology domain
PI3-kinase – Phosphatidylinositol 3′-kinase
***polh* –** Polyhedrin gene
SAD – Single wavelength anomalous dispersion
SH2 – Src homology 2 domain
SH3 – Src homology 3 domain
SPR – Surface Plasmon Resonance
Src –proto-oncogene tyrosine kinase
STB – Stirred-tank bioreactor
$TCID_{50}$ – Tissue culture infectious dose
$TCLD_{50}$ – Tissue culture lethal dose
TEC – Tyrosine kinase expressed in hepatocellular carcinoma
TH – TEC homology domain
T_m – Melting temperature
TNF – Tumor necrosis factor
TOH – Time of harvesting
TXK – Tyrosine-protein kinase
VEGFR2 – Vascular endothelial growth factor receptor

Chapter 1

1. Introduction

1.1. Prostate cancer

1.1.1. Epidemiology

Prostate cancer continues to be a major health threat and cause of death in men. According to Glocoban data, approximately 758 700 new cases and 142 000 deaths were estimated in 2012. The incidence rates are highest in the populations of North America, Northern and Western Europe, and Australia/New Zealand, where the practice and availability of prostate specific antigen testing and prostate biopsies have increased for asymptomatic patients.[1,2]

1.1.2. Pathology, genetics and risk factors

High-grade prostatic intraepithelial neoplasia is the main precursor lesion of invasive prostate cancer. The diagnosis of HGPIN denotes ca. 25% risk of detecting prostate carcinoma in subsequent biopsies, and the incidence may increase with aging.[3]

Cancer progression from a precancerous form to a more advanced carcinoma follows a characteristic pattern of genetic and epigenetic abnormalities, including the loss of tumor suppressor genes (e.g., PTEN, NKX3.1, p53, and Rb) and the overexpression of certain oncogenes (e.g., Myc, FCF8b, and CXCL3).[4–6] Androgen hormones (e.g., testosterone) are also intimately related with carcinogenesis.[7] Hence, therapies based on hormone suppression such as androgen-deprivation therapies (ADT) remain very effective against locally advanced and metastatic cancer, after localized therapies have failed. However, despite hormonal suppression, disease often progresses to an androgen-independent stage, also known as castrate-resistant prostate cancer (CRPC).[8]

Cancer resistance to chemotherapy through innate or acquired responses is one of the most major concerns in cancer therapy. The molecular and genetic heterogeneity of tumor cells and the complexity of the tumor microenvironment limit the effectiveness of current first-line chemotherapeutic drugs aimed at targeting a specific cellular component. Besides, these therapies carry severe risks and harsh side effects for patients due to the indiscriminate killing of both cancer and healthy cells.[9,10] Understanding the basis of prostate cancer resistance is crucial to develop improved active therapeutic molecules required to overcome drug resistance and lack of drug selectivity and effectiveness.[9,10]

1.2. Introducing Bone Marrow Tyrosine Kinase in chromosome X

Bone marrow tyrosine kinase in chromosome X (BMX) is a 78 kDa protein and a major member of the non-receptor TEC tyrosine kinase family. These include the tyrosine kinase expressed in hepatocellular carcinoma (TEC), Bruton's tyrosine kinase (BTK), interleukin (IL)-2-inducible T-cell kinase (ITK), and tyrosine-protein kinase (TXK).[11] BMX was first identified from a human prostate tumor xenograft CWR22 and its encoding gene is located on the human chromosome Xp22.2.[11–13] TEC family kinases present a very conserved core catalytic structure, composed of an amino-terminal pleckstrin homology (PH) domain, a TEC homology (TH) domain, a proto-oncogene tyrosine kinase (Src) homology 3 (SH3) and 2 (SH2) domains, and a carboxy-terminal catalytic tyrosine kinase domain. However, BMX lacks the proline-rich region in the TH domain and has a slightly different sequence in the SH3 domain (Figure 1.1).[12] These kinases are well known important regulators of several cellular processes including proliferation, differentiation, apoptosis and cell migration, and therefore they have received massive interest over the years as targets for anti-cancer therapies.[14]

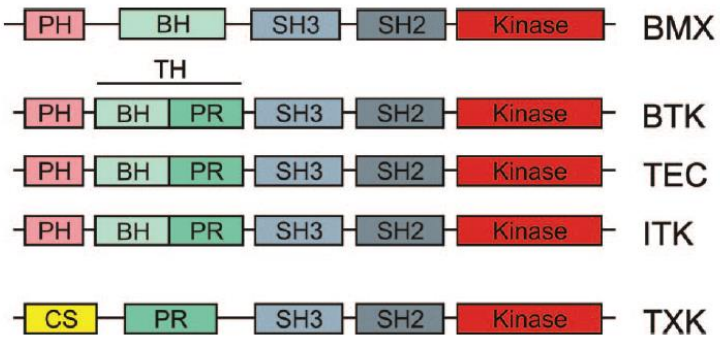

Figure 1.1. Sequence similarity of the TEC family members. PH, pleckstrin homology; CS, cystein string; BH, BTK homology; PR, proline rich; TH, TEC homology; SH, Src homology (Adapted from ref. 12).

1.2.1. Three-dimensional structure of BMX catalytic domain

BMX folds into the typical kinase bi-lobed core structure, with the ATP binding pocket located in a deep cleft situated between the carboxy- and the amino- lobes (Figure 1.2). The larger carboxy-terminal lobe consists mostly of α-helices, a few short β-strands, and it also includes a DFG (aspartic acid-phenylalanine-glycine) motif and an activation loop, located between the DFG and APE (alanine-proline-glutamic acid) motifs. Large conformational changes in this segment can switch kinase activity between its inactive and active form. The amino-terminal lobe is comprised of a five-stranded antiparallel β-sheet, one α-helix (the αC-helix), a G-loop (glycine-X-glycine-X-X-glycine loop), and a "hinge" region – a segment that connects both the carboxy- and the amino- lobes. The G-loop, the αC-helix, and the DFG-motif are the most flexible regions of the protein, meaning that the ATP binding site can assume different conformations in order to accommodate different shaped and sized ligands.[14–16]

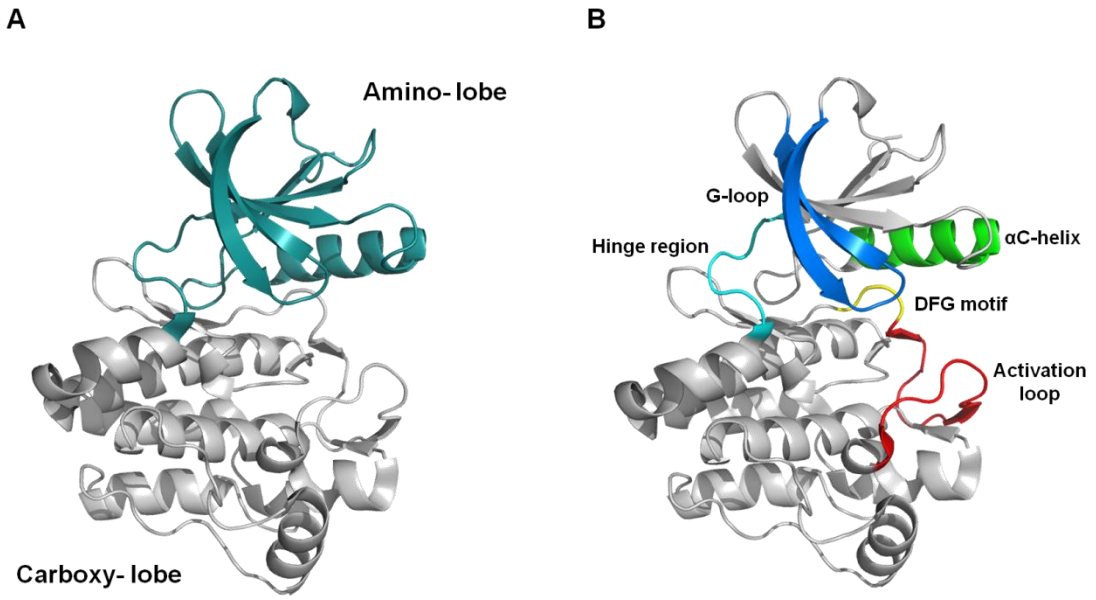

Figure 1.2. Structure of human BMX catalytic domain. A) The overall catalytic domain of hBMX tyrosine kinase with the amino- lobe colored in teal blue and the carboxy- lobe colored in grey. B) Important structural segments highlighted: the G-loop (marine blue); the hinge region (cyan blue); the αC-helix (green); the DFG motif (yellow); and the activation loop (red) (Adapted from ref. 15).

1.2.2. BMX expression and its role in prostate cancer

BMX is widely expressed in cells of the haematopoietic lineages, particularly in granulocytes and monocytes, as well as in the epithelial and endothelial cell lineages of the brain, prostate, lung, and heart.[12,17] BMX is also highly overexpressed in metastatic breast cancer and in prostate cancer, and has been implicated in tumorigenicity, motility, proliferation, angiogenesis, differentiation and apoptosis.[13,18–26]

BMX activation by Src family kinases in the plasma membrane requires an association of its pleckstrin homology (PH) domain either with phosphatidylinositol-3,4,5 triphosphate, a product of phosphatidylinositol 3′-kinase (PI3-kinase), or with the FERM domain of focal adhesion kinase (FAK).[13,20] PI3-kinase-mediated activation of BMX is induced by interleukin-6 (IL-6) and it has been implicated in neuroendocrine differentiation of prostate cancer cells.[13] BMX is also involved in prostate cancer cells migration and its activation is regulated by a FAK protein in response to extracellular matrix proteins.[20] In addition, the transactivation of BMX with the vascular endothelial growth factor receptor 2 (VEGFR2) is induced by the tumor necrosis factor (TNF), resulting in a direct activation of the P13K-Protein kinase B (Akt) angiogenic pathway.[19] More recently, BMX has been implicated in the adaptive compensatory mechanism of CRPC to androgen deprivation therapy.[22–24] BMX upregulation in CRPC induces tyrosine phosphorylation of the androgen receptor via its SH2 domain, enhancing androgen-receptor hormone-sensitivity and transcriptional activity, and consequently promoting prostate tumor growth and survival even in androgen-depleted conditions.[24] Furthermore, BMX overexpression in cancer cell lines suppresses a core component of the intrinsic apoptotic pathway through negative regulation of

the Bcl-2 homologous antagonist (BAK) function and inhibition of tumor suppressor protein p53, granting tumor cells the ability to escape apoptosis induced by chemotherapeutic drugs.[25,26]

BMX knockout mice remain fertile and have a normal life span, without presenting an obvious altered phenotype. This suggests a redundant role of BMX in cellular signal transduction, and therefore, therapies based on BMX inhibition might have limited side-effects.[27]

1.3. BMX kinase inhibitors

Full disclosure of BMX pathological role in carcinogenesis has led to an increased interest to design selective small inhibitors molecules to target BMX.[27] Liu and his colleagues have successfully synthesized a selective BMX inhibitor, called BMX-IN-1 (Figure 1.3). Selectivity is achieved since this kinase inhibitor is capable of forming an irreversible, covalent bond to the BMX active site, by reacting with a nucleophilic cysteine residue (496). As a result, the inhibitor irreversibly blocks the ATP binding pocket of BMX, thereby rendering the kinase inactive form. BMX-IN-1 inhibits also displayed moderate antiproliferative efficacies against prostate cancer cell lines at micromolar concentrations.[28]

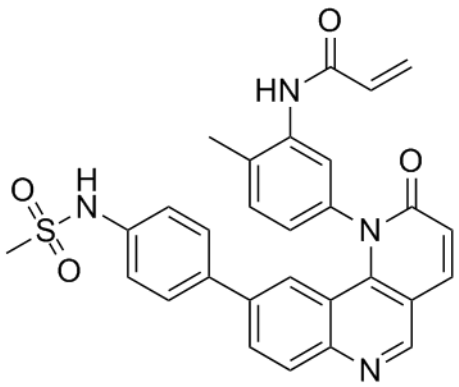

Figure 1.3. Chemical structure of BMX-IN-1.[28]

1.4. BMX production using the baculovirus-insect cell expression system

The baculovirus-insect cell expression system (IC-BEVS) has been proven successful for the production of many recombinant proteins.[30] TEC family kinases, such as BMX, BTK, and ITK have already been produced using this system.[15,31,32] IC-BEVS is based upon the ability to propagate a baculovirus shuffle vector (bacmid) in insect cells to deliver genes required for the expression of a target protein. The process of protein production via IC-BEVS entails two main steps: (1) the growth of insect cells to a desired concentration; (2) and the infection of insect cells with baculovirus to express the protein of interest during the exponential cell growth phase. The total amount of protein expressed will depend on the quantity of virus introduced into the cell culture.[30,33]

1.4.1. Baculovirus biology and life cycle: *in vivo* vs. *in vitro* cultures

Baculoviruses are enveloped, rod-shaped viruses with double-stranded DNA genomes, capable of replicating in the nucleus of an insect cell. They were initially used as ecologically friendly biopesticides to control insect pests, but soon after were also exploited for the production of eukaryotic proteins.[30,32] Recombinant baculoviral vectors derived from the *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) are highly effective in mediating gene expression in insect cells.[30] Since the majority of insect lines are susceptible to the AcMNPV infection, it is the most commonly used vector for producing recombinant proteins.[30]

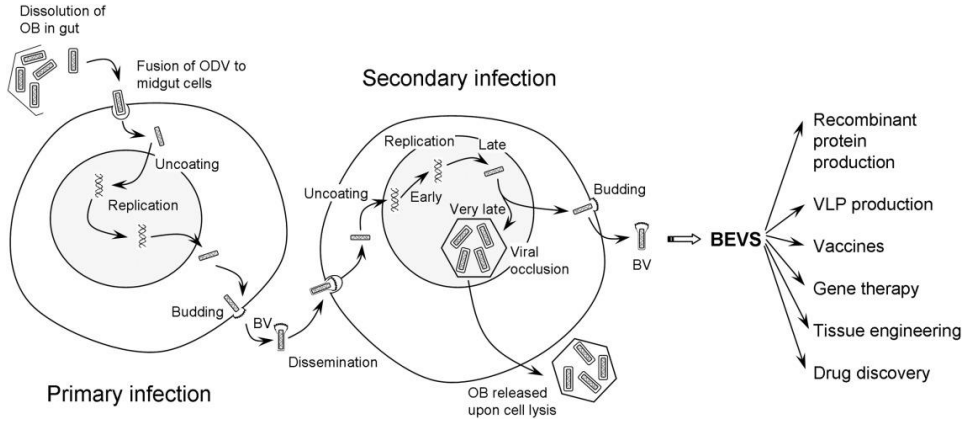

Figure 1.4. The replication cycle of a baculovirus.[30]

Baculoviruses have a biphasic life cycle in which two forms of virions are produced in different stages of infection and spread out through horizontal or vertical infection (Figure 1.4). Horizontal infection is mediated by the occlusion-derived virions (ODVs). These ODVs are occluded in a crystalline protein matrix made of polyhedrin, namely the occlusion bodies (OBs), which makes them resistant to the external environment conditions. The viral replication cycle begins when the OBs are ingested by the insect larvae. Once it reaches the midgut of the insect, the alkaline environment quickly dissolves the occlusion body, releasing its occlusion-derived virions. After passage through the peritrophic matrix barrier, the ODVs bind and fuse with the microvilli of columnar epithelial cells, in a process mediated by a protein complex present in the virion envelope, the PIF complex. Afterwards, the nucleocapsids are released into the cytoplasm and migrate to the nucleus to establish the infection.[30,34,35] The immediate/delayed early phases of infection initiate from 0 to 6 hours post infection (hpi) and are characterized by the expression of viral transregulators and genes implicated in establishing the infection, viral replication and host manipulation.[30] The late stage (6 to 18 hpi) is associated with the production of nucleocapsids. The transition from the early to the late stages of infection is marked by the onset of DNA replication and protein synthesis in the host cell. The newly assembled nucleocapsids are transported from the nucleus to the membrane for budding, originating the budded virions (BVs). These budded virions are responsible for the systemic

dissemination of the virus from within the insect (vertical infection), and consist of a single nucleocapsid envelope enclosed in a modified structure derived from the plasma membrane.[30] A second cycle of replication initiates when the BVs infect new cells in a secondary infection – a process mediated by the BV envelope glycoprotein, GP64, via clathrin-mediated endocytosis.[34] The process continues similarly to what was described for the primary infection, with the migration of the nucleocapsids to the nucleus and subsequent establishment of infection. This time, however, the OBs are produced in the very late stages of infection (18 hpi) and are released to the environment by cell lysis with consequent death of the host, in order to continue horizontal transmission of the infection.[30]

The *in vitro* life cycle of the baculovirus is very similar to what happens *in vivo*, with the exception that the cells are directly infected with the BVs. In fact, OBs are considered unnecessary for the virus survival in controlled and monitored *in vitro* environments. Therefore, the gene responsible for the expression of the polyhedrin protein that composes the envelope of the OBs is replaced by the gene of the protein of interest. Expression of the foreign gene is driven by the polyhedrin promoter, which is highly transcribed during the late stages of infection.[36]

1.4.2. Baculovirus-insect cell expression system

The IC-BEVS is a very efficient and rapid system to produce protein at high levels of expression. This is achieved by using an engineered virus (AcMNPV-based vector) to infect non-modified insect cells and transform them into protein-producing factories.[30,37] These AcMNPV-based vectors are circular double-stranded covalently closed DNA (usually 80 to 180 kbp in size) and contain multiple recognition sites for many restriction endonucleases. As a result, AcMNPV-based vectors carry many desired properties to facilitate the insertion and expression of one or multiple heterologous genes in insect cells for customized production work.[36,38,39]

The first cell line to be commonly used for AcMNPV propagation was the Sf-21 clone, developed from immature ovaries of pupal tissue of the armyworm insect *Spodoptera frugiperda*.[40] Nowadays, the use of Sf-21 cells has diminished in research and industrial fields, and instead they have been replaced by Sf-9 cells. Sf-9 cell lines are derived from Sf-21 and are capable of producing higher quantities of infectious baculovirus, due to their improved growth and infection characteristics. Both cell lines are well adapted to suspension cultivation; however, Sf-9 cells are less fragile and more tolerant to osmotic pressure and shear stress than Sf-21.[41]

Insect cells perform most of the oligomerization and posttranslational modifications (e.g., glycosylation, palmitoylation, myristoylation, fatty acid acylation, amidation) similarly to those found in mammalian cells. Still, certain modifications are not possible, since insect cells are not capable of generating complex N-glycans with terminal sialic acids. The differences between the desired and actual glycosylation patterns place some limitations regarding the utility of the recombinant glycol-proteins.[30,42] These limitations were tackled by engineering baculovirus to

express chaperone proteins, and by generating insect cells that express mammalian glycosylation patterns that were lacking in the parent cell line.[30,42]

1.4.3. Generating the recombinant bacmid in *E. coli* cells

Commercially available kits, such as the Bac-to-Bac® Expression System (Invitrogen, California, USA), enable the generation of the recombinant baculovirus (bacmid) based on AcMNPV vectors by site-specific transposition in competent *E. coli* cells. In this case, the gene of interest is first cloned into a donor plasmid by replacing a non-essential DNA gene, usually the polyhedrin (*polh*) (Figure 1.5). The competent *E. coli* containing the parent bacmid are transformed by site-specific transposition between the mini-Tn7 elements on the recombinant donor plasmid and the mini-attTn7 target site on the bacmid. Recombination of the parent bacmid occurs with the help of transposition proteins provided by the helper plasmid in the host bacteria. The baculovirus particles are generated once the resulting recombinant bacmid is introduced into insect cells.[36,43]

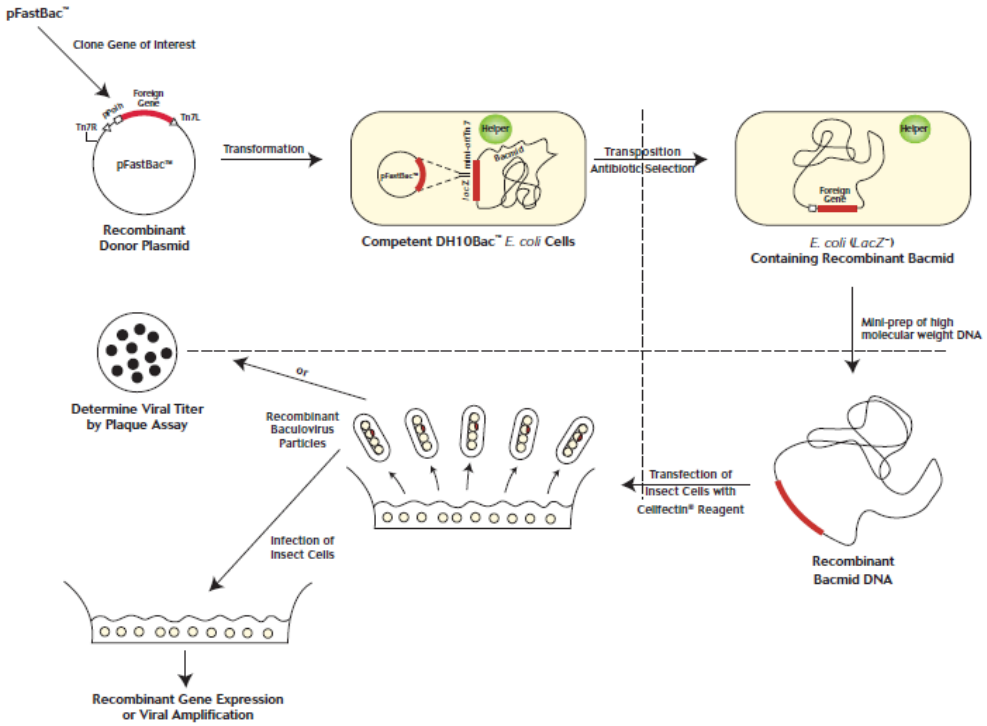

Figure 1.5. Generation of recombinant baculovirus bacmid and gene expression with the Bac-to-Bac® Expression System.[43]

The initial cloning of the target gene in *E. coli* cells allows a faster and more efficient generation of recombinant baculovirus in comparison with the standard homologous recombination in insect cells. Indeed, the construction and purification of the recombinant baculovirus vectors in insect cells would take as long as 4 to 6 weeks and would often result in low viral concentrations or titers.[36]

1.4.4. Bioprocess engineering aspects of insect cell culture

Within the different types of bioreactors, the stirred-tank bioreactor (STB) is the most commonly used to culture insect cells in robust, well characterized, and cost-effective processes. In this system, insect cells can be inoculated attached to solid surfaces, but normally they are directly initiated into suspension culture, without requiring further adaption to the cultivation media.[30] Besides, insect cells are metabolically less demanding than mammalian cells and do not accumulate as much ammonia or lactate, which can hinder culture performance.[44] Batch cultures are the most preferable modes of operation for large-scale production of protein via IC-BEVS. Batch cultures have the advantage of being simpler, more flexible and inherently suited to a process involving a lytic infection cycle.[30]

The main goal of a production process is to maximize the volumetric productivity of the protein of interest and enhance product quality. For this purpose, a set of critical parameters must be controlled. In addition to optimization of cell density and feeding strategies, other important parameters include culture temperature, the pH, the concentration of the dissolved oxygen and carbon dioxide, the multiplicity of infection (MOI), the time of infection and the time of harvesting (TOH).[30] The most important one is the MOI, since irreproducible MOIs can have a tremendous impact in protein productivity.[45,46] MOI is the quantity of plaque forming units (pfu) or infectious particles infecting a given cell population at the time of infection.[30] In theory, the infection with a MOI equal to 1 pfu/cell would result in each cell receiving an infectious particle; however, the infection process is merely hypothetical, meaning that the number of viral particles absorbed by each cell will vary. In fact, the probability that an individual cell will receive *n* viral particles when inoculated with an MOI of *m* can be modeled using a Poisson distribution:

$$P\{n\} = \frac{m^n e^{-m}}{n!} \qquad \text{Eq. (1)}$$

where *P{n}* is the probability for a cell to be infected with *n* infectious particles, and *m* is the MOI used at the time of infection.[47]

If by underestimation of the viral titer, higher MOIs are used at the time of infection, it can lead to a severe metabolic burden effect which will immediately affect cell growth, and consequently result in low protein productivity. Similarly, use of lower MOIs due to overestimation of viral titer, results in low protein production levels due to non-synchronous infection. Therefore, it is really important to minimize errors associated with MOI variability by selecting an appropriate titration method to accurately determine the titer or concentration of the infectious particles in solution. The most commonly applied methods for baculovirus titration are cell growth cessation assay (CGCA)[46] and microculture tetrazolium (MTT)[48]. Both methods rely on the cytopathic effects caused by the virus invasion in the host cell during the infection. The extension of this phenomenon is dependent on the viral dose and thus it can be correlated with viral titers.[45,46]

1.5. Sample quality assessment

1.5.1. Dynamic light scattering (DLS)

Dynamic light scattering (DLS), also known as photon correlation spectroscopy or quasi-elastic light scattering, is a technique used to measure the size distribution, stability and aggregation state of a protein. This technique measures the Brownian motion of particles in a liquid solution caused by the bombardment of molecules that surround them, and it correlates this motion with the size of particles.[49]

In DLS, when a monochromatic beam of laser light encounters the molecules in a sample, the light is scattered in all directions. Since particles are continuously changing their position due to the Brownian motion, this causes the intensity of the scattered light to fluctuate over time. These fluctuations give information about the velocity of diffusion of the molecules, and in turn the size of the macromolecules is determined: larger molecules tend to diffuse slowly, resulting in similar positions in different time points; as opposed to smaller molecules, which diffuse faster and therefore do not assume the same positions over time.[49]

Since the motion of the macromolecules in solution is dependent on their size, temperature and buffer composition, DLS can be used as a complementary technique to evaluate homogeneity and chemical purity of protein samples in different conditions.[50,51] X-ray crystallography is a technique used to determine the molecular structure of a target molecule and is highly dependent on the quality of the diffracted crystals. DLS can be helpful in understanding how the different variables influence the stability and the aggregation state of the protein in solution, and consequently, the likelihood of protein crystallization. Predominately monodisperse protein solutions are more likely to crystallize, while polydisperse solutions are more likely to result in poorly diffracting crystals or no crystals at all. This information can be used to change the buffer composition or to optimize experimental conditions in order to improve the protein behavior in solution.[51]

1.5.2. Circular Dichroism (CD)

Circular dichroism refers to the differential absorption between the two circular components of plane-polarized light, one rotating counter-clockwise (left-handed, L) and the other clockwise (right-handed, R).[52] Unequally absorption of the two components results in an elliptically polarized radiation, as shown in Figure 1.6, A-II. In contrast, none or equal absorption of the L and R components regenerates radiation polarized in the original plane (Figure 1.6, A-I). A CD instrument, such as a spectropolarimeter, can detect this differential absorption between the L and R circularly polarized components and report it in terms of ellipticity (θ) in degrees as a function of the wavelength (λ). [52,53]

A CD signal is generated when a chromophore is chiral (optically active) either because of its structure (e.g., a carbon atom with four different substituents), it is covalently linked to a chiral centre or it is placed in an asymmetrical environment provided by the three-dimensional

structure of a molecule.[53] Proteins constitute very asymmetrical environments and therefore they can be analyzed by CD to obtain complementary structural information from several spectral regions. The type of information obtained from protein CD spectra includes the content of α-helices, β-sheets, and turns, and also the state of denaturation (Figure 1.6, B). CD can also be applied to monitor protein conformation changes due to temperature, mutations, denaturants or binding ligands.[54]

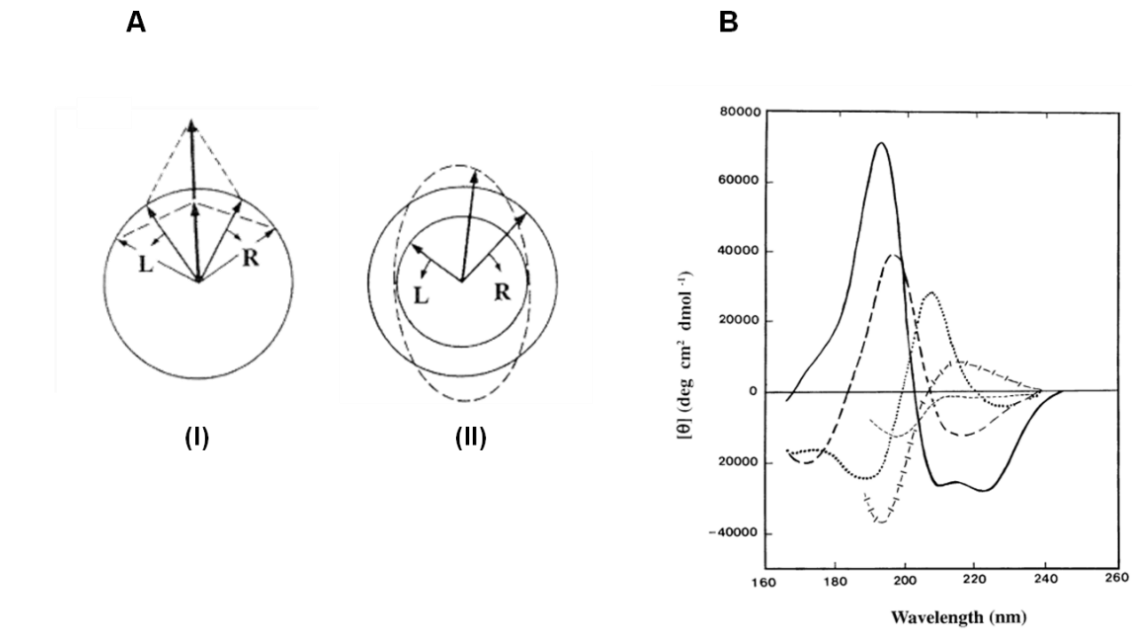

Figure 1.6. Origin of the circular dichroism effect and protein CD spectra. (A) The left (L) and (R) components of plane polarized light: (I) the two components have the same amplitude and the combination regenerates a plane polarized plane; (II) the two components have different magnitudes and the combination of the two results in an elliptically polarized light. (B) Far UV spectra for different types of secondary elements of proteins. Solid line, α-helix; long dashed lined, anti-parallel β-sheet; dotted line, type I β-turn; cross dashed line, extended 3_1-helix or poly (Pro) II helix; short dashed line, irregular structure (Adapted from ref. 52).

A number of algorithms exist which use the far UV CD spectra to provide an estimation of the secondary structure and folding properties of an unknown protein molecule. Widely used algorithms, such as k2d, analyze entered data by comparing it with source datasets comprising the CD spectra of proteins whose structures have been already solved by X-ray crystallography.[52]

1.6. Protein-ligand complex characterization

1.6.1. Differential scanning fluorimetry (DSF)

Differential scanning fluorimetry or fluorescence-based thermal shift assay (ThermoFluor) is a technique used to study the thermal stability of proteins under various conditions.[55] Protein stability is determined by gradually increasing the temperature and measuring the activation of a fluorescent dye that binds to the hydrophobic parts of the protein as it unfolds.[55,56]

This process is quantifiable through the midpoint of the thermal unfolding curve, which corresponds to the melting temperature (T_m) of a protein. At this point, both the folded and unfolded states are present in equal concentrations (Figure 1.7). A positive T_m shift under different conditions is indicative of an increase in protein stability. In theory, nearly all ingredients in solution, buffers, salts, detergents, or even ligands can establish interactions with several protein residues, thus increasing its structural stability and T_m.[55]

DSF can be designed and implemented to perform high-throughput screening of conditions that better stabilize proteins, and minimize denaturation and protein aggregation, without requiring large amounts of protein. This technique can also be applied to identify low-molecular-weight ligands that bind and stabilize proteins.[55,57] In X-ray crystallography, DSF is very commonly used to test a number of factors that can influence crystal growth, such as buffers, pH, ionic strength, additives, precipitants, or ligands, with the purpose of finding conditions leading to crystallization.[58]

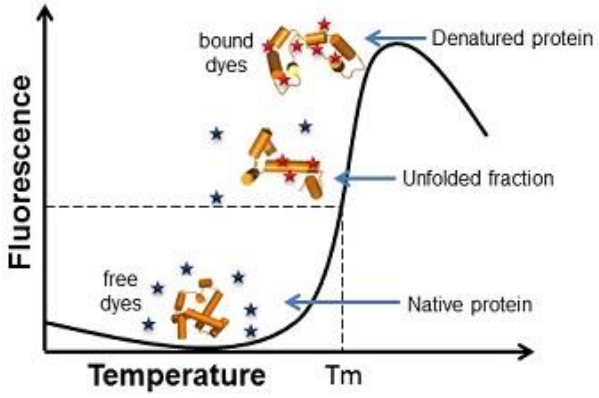

Figure 1.7. Representation of a protein melting curve in the presence of a fluorescent dye. The fluorescence intensity is plotted as a function of the temperature for the unfolding protein.[59]

1.6.2. Surface Plasmon Resonance (SPR)

Surface Plasmon resonance (SPR) spectroscopy is a label-free detection method to study biomolecular interactions in real-time with high-sensitivity.

In SPR, a receptor molecule is immobilized on a thin metallic surface (typically a gold surface) and kept in contact with a solution. As light is shone through a prism at the surface, it excites the electrons present in the gold surface, causing them to resonate (Figure 1.8).[60,61]

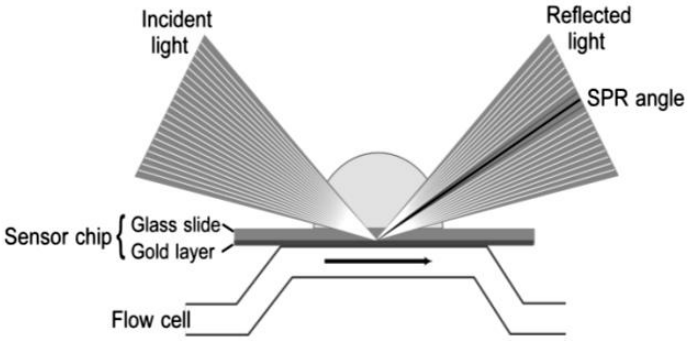

Figure 1.8. Principle of a SPR detection experiment.[62]

These resonating electrons are also known as surface plasmons. The defined angle at which resonance occurs depends not only on the metal, but also on the refractive index of the medium just above the gold layer. Consequently, when a solution of analyte molecules is flowed over the immobilized molecule, molecular-binding events occur between the receptor molecule and the analyte, inducing an increase in the refractive index, which can be directly monitored at the SPR sensor surface or sensorgram (Figure 1.9). Once the saturation limit is reached, the analyte is progressively removed from the continuous flow by buffer washing, resulting in a decrease of the response in the sensorgram. Hence, SPR can be employed to evaluate the kinetics and affinity of an interaction between a protein and inhibitor compound at specific given conditions. From this evaluation, both the rate of association (k_{on}) and dissociation (k_{off}) can be determined. The binding affinity of the ligand to the receptor molecule is estimated from the ratio of the rate constants ($k_D = k_{off}/k_{on}$).[60–62]

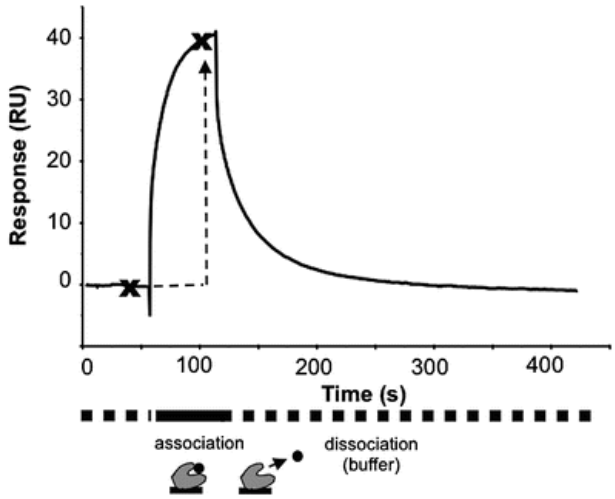

Figure 1.9. A typical binding cycle observed with an SPR sensorgram. The dotted lines represent the running buffer, while the solid line indicates the injection of an analyte, resulting in the binding of the analyte to a receptor molecule immobilized on the chip surface. [63]

1.7. X-ray crystallography for protein-ligand structure determination

X-ray crystallography is a technique used to determine the three-dimensional structure of a protein from the diffracted X-ray beams. The diffraction phenomenon takes place since both the wavelength of the incident X-ray beam and the interatomic distances in the crystal are of the same order of magnitude (10^{-10} m). By using this type of electromagnetic radiation, structural information can be provided at an atomic resolution up to 0.5 Å. To determine 3D structures by X-ray crystallography, the process entails the growth of high-quality protein crystals, measurement and analysis of the X-ray diffraction pattern, calculation of an electron density map and its interpretation to build an atomic model of the protein.[64]

1.7.1. Protein crystallization

Crystals are 3-dimensionally ordered entities of repeated molecules, stabilized by non-covalent interactions and solvent channels. They are grown by reducing the protein solubility in aqueous solution under specific physiochemical conditions.[64] Crystals are characterized by the repeated translations along the three special directions of a fundamental building block - the unit cell. A unit cell is a parallelepiped defined by the lengths of three intersecting, non-parallel edges (a, b and c), and by the three angles between them (α, β, and γ). There are only 14 different lattice types (Bravais lattices) belonging to seven possible crystal systems. Also, there are 32 point groups that represent all of the possible unique combinations of crystallographic symmetry elements (2,3,4,6-fold rotation axes and mirror planes). The combination between the 32 point groups and the 14 Bravais Lattices results in 230 unique Space Groups, which describe the only ways in which identical objects (molecules) may be arranged in an infinite lattice. The asymmetric unit is contained within the unit cell and represents the smallest portion of a crystal structure that can be transformed to generate a complete unit cell by using only symmetry operations allowed by Space Group.[64–66] The repetition of the unit cell throughout the three-dimensional space allows the determination of the position of each atom in a unit cell by means of its diffraction coordinates (Figure 1.10).[64]

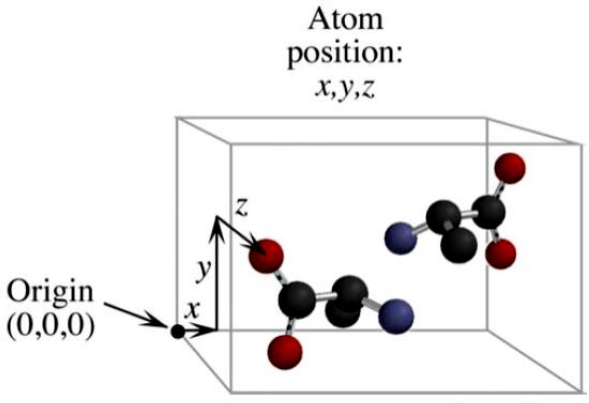

Figure 1.10. Representation of one unit cell. The position of an atom in the unit cell determines a specific set of spatial coordinates x, y, z. [64]

Protein crystal formation is a highly directional and organized process which requires rigorous interactions between molecules.[67] In theory, such molecular interactions can be achieved by bringing the protein solution to a supersaturation zone, i.e., by reducing the protein solubility either by altering the pH or by adding salts, polymers or ions.[64,67] Supersaturation is required to break the energy barrier that exists when forming small tiny crystals – a phenomenon also known as nucleation.[67,68] Bringing a protein solution into a supersaturation state can be achieved through several techniques: vapor diffusion, batch, dialysis and free interface diffusion. The most common employed one is vapor diffusion and it can be performed through hanging or sitting-drop. In the vapor diffusion method, a small drop containing protein, buffer and precipitant, is equilibrated with a precipitant in higher concentrations within a closed system. As the water vapor from the drop starts to diffuse and condense in the reservoir due to the different precipitant concentrations in drop and reservoir solutions, the protein in the drop is gradually driven into supersaturation. If the ideal conditions for nucleation are reached, protein molecules start to organize into microscopic ordered aggregates to produce small crystal nuclei. Once the nuclei are formed, the protein concentration in solution gradually decreases, driving the system into the metastable zone, where crystal growth occurs slowly.[64,69] Figure 1.11 illustrates a complete crystal growth process through the vapor diffusion method in a two-dimensional phase diagram, as a function of precipitant and protein concentrations.

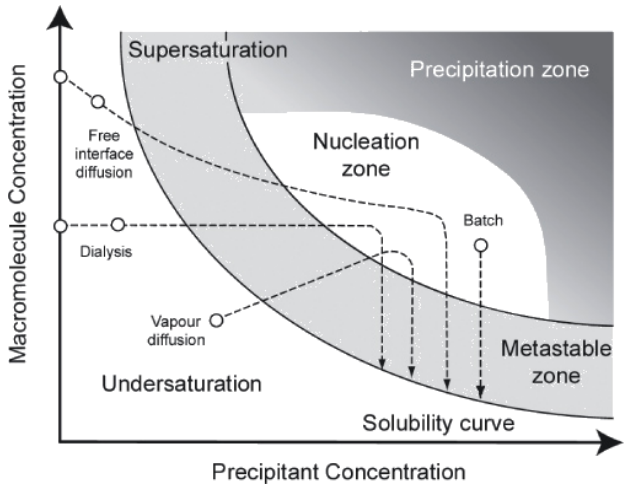

Figure 1.11. Representation of a protein crystallization phase diagram indicating different methodologies to reach the nucleation and metastable zones.[70]

The phase diagram is divided into different regions: a very high supersaturation ("precipitation"), where the excessive protein molecules in solution form amorphous aggregates; a nucleation zone, where the formation of small nuclei occurs; lower saturation ("metastable zone"), where only crystal growth is supported; and undersaturation, where the protein molecules are evenly dissolved throughout the solvent and no crystallization can occur.[67]

In reality, the chemical space in crystallization experiments is multidimensional rather than two-dimensional, embodying many physiochemical or biochemical variables which impact the thermodynamics of nucleation and crystal growth.[67] These include the chemical and

conformational purity of the protein sample, temperature, solution pH, ionic strength, rate of equilibration, type of precipitant used, protein and precipitant concentrations, and also the presence of ligands. As a result, the success of this process lies in controlling and testing a large number of conditions in order to obtain a reproducible source of diffracting crystals.[64,67]

High throughput methods coupled with liquid-handling robots (e.g., Mosquito® LCP crystallization robot) help to cover a broad range of parameters required to explore all the desired conditions for crystallization, while maintaining the efficiency and accuracy of an automated system. In addition, robotic crystallization experiments do not require large amounts of precipitant or protein solutions. Currently, there are numerous commercials kits for high throughput screenings available to explore various conditions for crystal growth.[69]

1.7.1.1. Crystallization of protein-ligand complexes

As previously described, protein crystals are a periodic molecular arrangement stabilized by weak non-covalent interactions, permeated by large solvent channels. Protein crystals typically contain from 30 – 80% of solvent in these channels that are in direct contact with the solution bathing the crystal, allowing small ligands to access and bind to the protein molecules in the crystal.[64,71] Compound incorporation in protein crystals can be achieved by co-crystallization and soaking methods.

In the first method, the compound is incubated with the protein prior to the crystallization experiment. The compound can either be added directly to the drop or pre-incubated with the protein. Co-crystallization is often applied when the compounds are not very soluble in aqueous solvents or the protein aggregates easily.[71,72]

The soaking method entails the incubation of a preformed crystal of the target protein (apocrystal form) with the compound. It can also be used to replace a co-crystallized inhibitor with a second one.[71] Soaking it is often the method of choice to obtain protein-ligand complexes, since it allows the preparation of many such complexes with different ligands from high-quality apocrystals. However, there are some disadvantages to be considered. For instance, the crystals may be sensitive to the solvent used for inhibitor solubilisation, requiring soaking in a stabilization buffer or cross-linking, e.g., with glutaraldehyde. Optimization of the soaking time and inhibitor concentration might also be required.[72,73]

1.7.2. X-ray sources and data collection

Having obtained a high-quality crystal, the next step is the recording and analysis of its X-ray diffraction pattern. To do so, the crystal is placed on a goniometer head and exposed to a monochromatic, intense X-ray beam. The diffracted beams are collected on a detector, generating a pattern from which the crystal symmetry (crystal system, lattice type and possible space groups) can be inferred (Figure 1.12). From the diffraction data, an image of the electron

density of the atoms in the crystal can be obtained and interpreted to build a three-dimensional structural model of the protein.[64]

Trial exposures can be performed *in house* or at synchrotrons. *In house* systems typically employ X-ray tubes to generate X-rays. In this case, a cathode filament is heated, releasing electrons that are accelerated by an electric field towards a metal anode (usually copper, molybdenum or chromium). The moving particles collide with the target atoms, undergoing deceleration and losing their kinetic energy which is converted to X-ray photons. At specific energy values that depend on the nature of the metal target, electrons from the innermost (K) shell can be ejected from the metal atoms. Then, an electron from a higher shell can drop into the K shell to fill the vacancy, and the energy difference between atomic orbitals is released as an X-ray photon with a characteristic energy. The *in house* X-ray sources use mirror systems or monochromators to select these X-ray photons for crystal exposure [64,74]

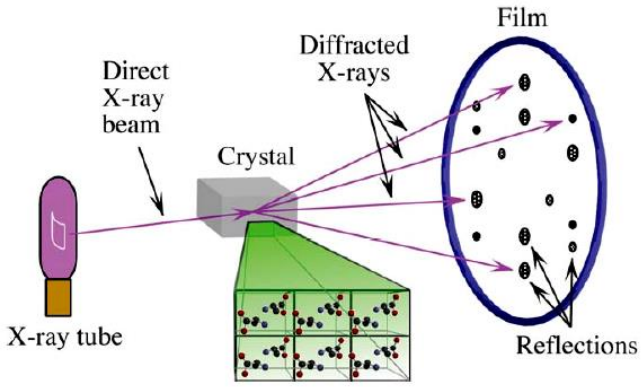

Figure 1.12. Diagram of an X-ray diffraction experiment. An X-ray beam generated by an X-ray tube is directed at a protein crystal. The result is a diffraction pattern obtained from X-ray scattering that is recorded on a detector (film). The intensities and positions of each reflection contain the information required to reconstruct the electron density map.[64]

Synchrotron X-rays are produced by electrons accelerated in a closed circular path forced by strong magnetic fields (Figure 1.13). The high energy electrons are directed to insertion devices (undulators or wigglers) located in particle storage rings. Dipole magnets present in undulators cause the traversing electron to oscillate, producing tangential X-ray photons that are captured by beamlines. Synchrotron radiation has several advantages when compared with *in house* sources; it combines extremely high intensity, high collimation, and tunability, to reduce the amount of time required for complete collection of data set and to study atomic features at lower resolutions. [64,74]

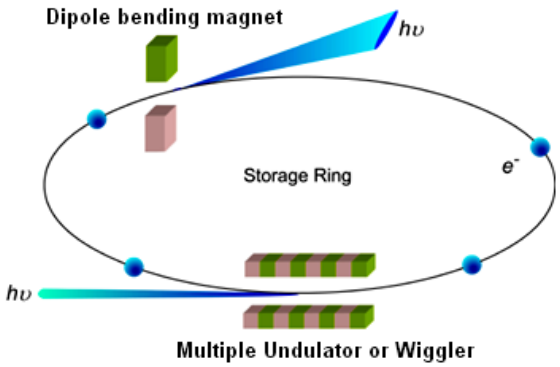

Figure 1.13. Schematic representation of a synchrotron (Adapted from ref. 74).

The diffraction pattern arising from the X-ray experiment is recorded on a detector behind the crystal. The geometrical arrangement of the diffracted beams (reflections) can be described by a reciprocal space lattice related to the crystal lattice. Each reflection corresponds to a reciprocal lattice point and can be geometrically characterized by three integer numbers h, k, l, also known as the Miller indices. According to Bragg's diffraction model, X-ray diffraction is a manifestation of the wave nature of electromagnetic radiation, and results from constructive interference between X-rays reflected from a family of parallel planes in the crystal, with interplanar spacing d. Constructive interference arises only for specific incident angles and depends on the interplanar spacing and the wavelength of the incident radiation (Figure 1.14).[64]

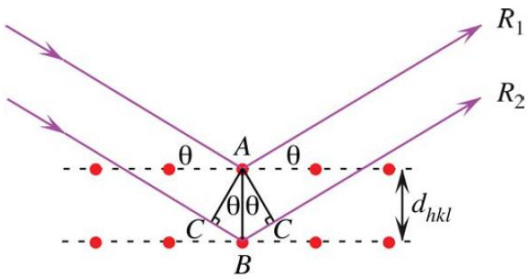

Figure 1.14. Schematic representation of Bragg's diffraction model. The ray planes R1 and R2 interfere constructively if the additional distance of R2 (d_{hkl}) is an integral number of wavelengths: $2d \sin \theta = n \lambda$.[64]

The different plane families intersecting different regions of electron density in the crystal, give rise to diffracted beams with different intensities (I_{hkl}), which carry valuable information for structure determination.[64]

1.7.3. Protein structure determination

The recorded intensities on the diffraction pattern, corresponding to the information determined by the arrangement of molecules within the crystal lattice, are converted into a three-dimensional model of the electron density map. The mathematical operation which describes this conversion is known as the Fourier transforms.

The electron density ρ at each point x,y,z of the crystal unit cell is given by Eq. (2), where V is the unit cell volume, $|\mathbf{F}_{hkl}|$ is a structure factor amplitude and ϕ_{hkl} its corresponding phase angle.[64]

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l |\boldsymbol{F_{hkl}}| \exp[-2\pi i(hx + ky + lz - \boldsymbol{\phi_{hkl}})] \quad \text{Eq. (2)}$$

The $|\mathbf{F}_{hkl}|$ values can be calculated from the measured diffraction intensities, however the phase angle information is lost. The summation runs over all the h,k,l indices of the measured diffraction intensities. However, since both the amplitudes and phases are required to obtain the electron density map and to build the structural model, the collected data are incomplete to reconstruct the real space information of the unit cell in the crystal. This impossibility of recovering the phase information originates the so-called "phase problem" in crystallography.[64] A range of techniques have been developed to overcome this problem: multiple isomorphous replacement (MIR), single or multiple wavelength anomalous dispersions (SAD/MAD), and molecular replacement (MR). MIR consists in preparing heavy-atom (HA) derivative crystals through soaking or co-crystallization. The addition of the heavy atoms introduces small changes the measured diffraction intensities and, if the native and derivative crystals are isomorphous (i.e., their unit cell dimensions are not significantly different) these isomorphous differences can be used to locate the HA substructure, thus allowing the estimation of the native phases. Similarly, the SAD or MAD methods utilize the intrinsic anomalous scattering properties of native metals such as iron or copper, sulfur atoms in cysteine residues, or selenium atoms in selenomethionine residues, to first determine their location in the unit cell and subsequently estimate phase angles. Finally, the MR method relies upon the existence of a known structure model homologous to the protein of interest. Usually, a high degree of amino acid sequence identity and similarity is taken as an indication of a similar 3D structure, and the known structure is used as a phasing model to estimate the initial phases of the unknown structure and calculate an electron density map to build an atomic model of the protein of interest.[64,75]

1.7.4. Refinement and cross-validation

Once a good starting model is obtained, the structural parameters describing the model can be optimized through a series of refinement cycles to improve agreement with the observed data. The initial model is submitted to manual rebuilding by adjusting the atomic positions to the computed electron density. Better phases can be calculated from the adjusted coordinates, allowing to redetermine the electron density map with a higher accuracy. In each cycle of refinement, new changes can be applied to the structural model, for instance: introduction of missing atoms, redefinition of atomic coordinates (x, y, z), or substitution of a wrong atom type. Also, because the detail of the electron density maps of protein crystal structures are seldom detailed enough to allow the visualization of individual atoms, stereochemical dictionaries specifying the geometrical parameters of amino acids and of any ligands and protein-metal bonds present in the structure must be included in the refinement so that a structural model with

chemical significance can be obtained. This process continues until the agreement between the calculated $|F_{calc}|$ and observed $|F_{obs}|$ structure-factor amplitudes, is maximized (Eq. (3)).

$$R = \frac{\sum||F_{obs}| - |F_{cal}||}{\sum |F_{obs}|} \qquad \text{Eq. (3)}$$

The R-factor is a measure of quality of the atomic model obtained from the experimental data. Cross-validation of the model is achieved by calculating a free R-factor, which is used to guard against possible misinterpretation of the experimental data or overfitting. The free R-factor is calculated from a small subset (5 to 10%) of the experimental data that is not used in the refinement calculations. The free R-factor is always higher than the R-factor, sometimes by as much as 10%, but the refinement protocol should result in the smallest possible difference. Differences larger than 10% indicate problems with the model that should be carefully investigated. The quality of the model is also validated by other parameters, such as Ramachadran plot outliers, side-chain conformations and root-mean-square deviation (r.m.s.d.) of the geometrical parameters with their dictionary values.[64]

1.8. Scope of the thesis

The realization of the full impact of the role played by BMX in prostate cancer has encouraged the design of more potent and selective inhibitors with improved pharmacological properties to overcome the current limited therapies. Thus, the Gonçalo Bernardes group located at the iMM, has developed a series of irreversible BMX inhibitors, analogues of BMX-IN-1[28], which have shown an increased inhibitory capacity when compared to BMX-IN-1. Since the molecular basis underlying BMX inhibition is still not fully understood, further elucidation of the functional and structural characteristics of BMX kinase alone or in complex with these irreversible inhibitors is necessary for feasibility testing of therapeutic efficacy. X-ray crystallography can also provide additional information regarding a molecular view of the BMX ATP binding pocket and how it binds to different inhibitors.

To fulfill the requirements, the work of this master thesis aimed at producing recombinant human BMX through IC-BEVS, characterizing biochemically and biophysically the protein alone or in complex with different inhibitors, and undertaking crystallization trials to determine the structure of the protein with BMX-IN-1 analogues.

For the first part of the thesis, the human BMX gene was cloned into a baculovirus shuffle vector in competent *E. coli* and expression of the target protein was carried out by transfecting Sf-9 insect cells in order to produce stock recombinant baculovirus particles. Viral stock concentration was determined by the CGCA and MTT titration methods, and viruses were used to infect insect cells in stirred-tank bioreactors for large-scale BMX production. Next, different biochemical and physical approaches such as SDS-PAGE, Western Blot, Dynamic Light

Scattering, and Circular Dichroism were carried out to evaluate the quality of the protein sample for crystallization, and Differential Scanning Fluorimetry and Surface Plasmon Resonance methods were used to characterize BMX in complex with different inhibitors. In a final step, unfrozen 6xHis-tagged BMX was co-crystallized with potent covalent inhibitors and the purification and crystallization experiments were optimized in an attempt to obtain high-quality crystals for X-ray diffraction.

This thesis provides substantial information regarding the biochemical and biophysical characterization of the human recombinant BMX expressed via IC-BEVS and further explores the inhibitor capabilities of covalent inhibitors based on BMX-IN-1. Hopefully, these findings will encourage further studies aiming to develop therapeutically active molecules in order to improve clinical outcomes for prostate cancer patients.

Chapter 2

2. Methodology

2.1. Synthesis of the recombinant donor plasmid

The human BMX tyrosine kinase gene was synthesized and assembled into a donor plasmid (pFastBac™ vector) by Invitrogen GeneArt™ gene synthesis. The guidelines for the plasmid construct are described in ref. 15. The insert included an N-terminal 6xHis-tag, a tobacco etch virus (TEV) protease cleavage site, and the gene sequence of human BMX catalytic domain [His_6-TEV-hBMX (residues 411-675)]. The recombinant donor plasmid also contained two antibiotic resistance genes (ampicillin and gentamicin), two cloning sites of XhoI and HindIII restriction enzymes, and a *polh* promoter for high-level gene expression in insect cells (Figure 2.1).

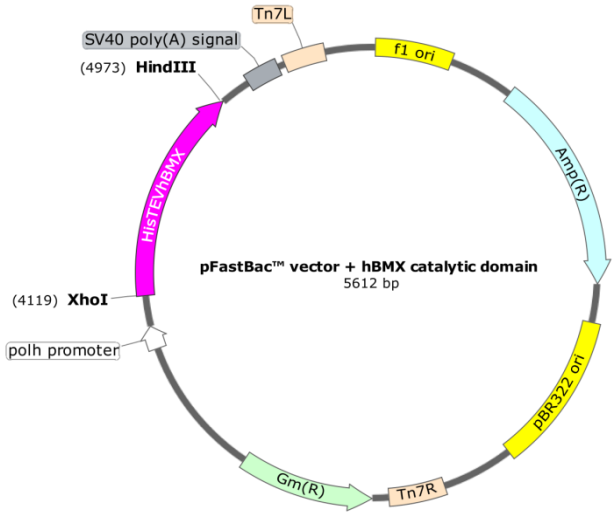

Figure 2.1. Recombinant donor plasmid map containing the hBMX gene sequence (provided by the Invitrogen GeneArt™).

2.2. Transformation of competent *E. coli* cells

Competent DH10Bac™ *E. coli* cells were transformed with 50 ng of recombinant donor plasmid. After being exposed to a heat shock, 300 µL of SOC media were added to the mixture and then incubated for 8 hours on a shaker at 200 rpm, 37ºC, to ensure the transposition of the target gene from the donor plasmid into the parent bacmid in the host cell. The content of each transformation was concentrated to 1x and 100x times and plated on an LB agar plate containing 50 µg/mL kanamycin, 7 µg/mL gentamicin, 10 µg/mL tetracycline, 100 µg/mL X-gal, and 40 µg/mL IPTG. The plates were incubated at 37ºC for 72 hours to enable the appearance of LacZα gene disrupted white colonies. The recombinant bacmid DNA was isolated from a selected white colony and purified with ZR BAC DNA Miniprep Kit (Bio-Rad, California, USA).

2.2.1. Analysis of the recombinant bacmid DNA by PCR

To confirm the correct transposition of the hBMX gene into the bacmid in the host cell, purified bacmid DNA from a white colony was analyzed by PCR. Amplification of the plasmid was performed following the guidelines of DNA *Taq* polymerase (Thermo Fisher Scientific, Massachusetts, EUA) PCR protocol (Table 2.1). A negative control was also performed by extracting the plasmid DNA from a blue colony. The final PCR product was expected to have 1311 bp and analysis was carried out on a 1% agarose gel.

Table 2.1. Primers sequence and PCR reaction conditions.

Preparation	**PCR cycling conditions**			
Total reaction volume of 50 µL:	**Steps**	**Temperature**	**Time**	**Cycles**
1) 1X *Taq* Buffer 2) 0.2 mM dNTP Mix	Initial denaturation	94ºC	3 min	1
3) 0.5 µM forward and reverse primers	Denaturation	94ºC	45 sec	
4) 3 mM $MgCl_2$	Annealing	52ºC	45 sec	35
5) 1-2 µL template DNA	Extension	72ºC	4 min	
6) 1.25U *Taq* DNA Polymerase 7) Water nuclease-free to 50 µL	Final extension	72ºC	10 min	1
a) PCR product 1311 bp **b) Primers:** Forward: 5′-GAGAACCTGTACTTCCAAGGC-3′ Reverse: 5′-TGTGGGCGGACAAAATAGTTG-3′				

2.3. Baculovirus-insect cell expression vector system

2.3.1. Insect cell culture

Spodoptera frugiperda derived Sf-9 cell lines were obtained from the European Collection of Cell Cultures (No. 89070101, ECCAC) and cultivated in suspension in serum-free media SF900II (Gibco, Glascow, UK) at 27°C in 500 mL shake flasks (50 mL working volume), and stirred at 100 rpm. The cell cultures were diluted routinely every 3-4 days using an inoculum of $0.8x10^6$ cells/mL. Cell density, percentage of viability and size were estimated by Cedex HiRes Analyser (Roche Diagnostics, Mannheim, Germany).

2.3.2. Insect cell transfection

The recombinant bacmid was introduced into Sf-9 cells by cationic liposome-mediated transfection using Cellfectin™ II reagent (Invitrogen, Life Technologies). Briefly, cells were seeded at $0.8x10^6$ cells per well in 6 well plates (Sigma-Aldrich, Darmstadt, Germany), with 2 mL of Grace media not supplemented (Gibco), and then allowed to settle for 15 minutes at room temperature. Both the Cellfectin reagent and the purified recombinant bacmid DNA (2 µg) were diluted in 100 µL of Grace medium not supplemented, mixed together, and then added dropwise onto the cells. After an incubation for 5 hours at 27°C, the DNA-lipid mixture (supernatant) was

removed and replaced with 2 mL of Grace's Insect supplemented medium with 10% FBS (Gibco). The cells were incubated at 27°C for 72 hours and monitored for signs of viral infection. The first generation baculoviruses (P1) were harvested by centrifugation first at 500 g, for 5 minutes at 4°C, then once more at 2000 g, for 15 minutes at 4°C. Virus stock was stored at 4°C, protected from light.

2.3.3. Amplification of the viral stock

The viral stock obtained from transfecting Sf-9 cells is a low-titer viral stock. The amplification of the initial viral stock is often required to generate higher titers and for optimal results in protein expression.[76] To amplify the P1 baculovirus, cells were seeded at $1x10^6$ cells/mL in 500 mL shake flasks (50 mL working volume). The titer of the P1 viral stock was not determined; instead we assumed both the possibility of having a small or higher titer, and infected the cells with 100 µL or with 500 µL of the same stock. P2 baculoviruses were harvested by centrifugation after the viability percentage dropped to values ranging between 70 and 90%, corresponding to 144-168 hours post-infection (hpi). For P2 baculovirus amplification, cells were infected at a low MOI of 0.01 pfu/cell, as recommended by the Bac-to-Bac® Expression System manual (ref. 43). Baculoviruses were harvested by centrifugation at 80-90% of cell viability, approximately 72-96 hpi. The baculoviral stocks were stored at 4°C, protected from light.

2.3.4. Virus titration by CGCA and MTT methods

Following the methodology described in Roldão *et al*., 2009 (ref. 46), Sf-9 cells were seeded at $1x10^6$ cells/mL, in 100 mL shake flasks (13 mL of working volume), and infected with a series of virus serial dilutions ranging from 10^{-1} to 10^{-4} times. A negative control consisting of non-infected cells was also performed. Non-treated data was analyzed using the four-parameter logistic regression. Data is fitted to a nonlinear regression curve built into maximum and minimum values, an estimated tissue culture infectious dose ($TCID_{50}$ – determined by the solver function), and a slope factor. Viral titers (pfu/mL) were determined by multiplying the ratio between seeding density and $TCID_{50}$ by the conversion factor 0.69.

The titration of the baculovirus stock using the MTT methodology was performed as described in ref. 46. Using a 96-well tissue culture plates (Sarstedt, Newton, USA), 10 µL of 5 x 10^5 cells/mL of Sf-9 cells in SF900 II medium (Gibco) were seeded in each well and allowed to settle for 2 hours at 27ºC. After incubation, the supernatants were removed and 150 µL of viral stock solution diluted from 10^{-1} to 10^{-10} times were added per well. Plates were incubated at 27ºC for 6 days. Thiazolyl blue tetrazolium bromide (Sigma Aldrich, St. Louis, USA) was added at a volume of 10 µL/well to a final concentration of 5 mg/mL. Plates were incubated at 27ºC for 4 hours. The supernatants were removed and 150 µL dimethyl sulfoxide (Sigma-Aldrich) were added to each well to ensure total solubility of formazan crystals. Plates were placed on a

shaker with soft agitation for 10-20 minutes. Absorbance readings (570 and 690 nm) were carried out on a SPECTRAmax™ microplate reader (Molecular Devices Corporation, Sunnyvale, USA). Raw data was analyzed by using Prism for Windows 6.01 (GraphPad software, La Jolla, USA) and fitted to a dose-response curve to determine EC_{50} (dilution that provokes a response halfway between maximum and baseline dilution). The tissue culture lethal dose 50 ($TCLD_{50}$) and viral titers (pfu/mL) were determined from EC_{50} by using the mathematical regressions described in ref. 46.

2.4. hBMX production

2.4.1. Small scale feasibility test and quality assessment

Sf-9 cells seeded at 1 x 10^6 cells/mL in 100 mL shake flasks (13 mL of working volume) were infected with different MOIs: 1, 0.1, 0.01, and 0.001 pfu/cell. Cultures were maintained at 27°C and stirred at 100 rpm, until cell viability dropped to 70-90%. Culture samples were taken daily and centrifuged at 500 g for 5 minutes. Harvested cells were rinsed with 100 µL PBS and kept at -80ºC.

To release the intracellular hBMX, the pellet samples were lysed with SDS-PAGE buffer (2X), followed by denaturation at 95ºC for 5 minutes. SDS-PAGE gel was carried out using 4% stacking and 15% separating gels. Protein bands were visualized by Coomassie Blue staining or by performing a Western Blot. For the Western Blot, bands were transferred to an Immun-Blot® PVDF Membrane (Bio-Rad Laboratories). The membrane was rinsed in blocking buffer (5% non-fat milk in TBS with 0.05% Tween) and left at room temperature for 1 hour, with soft rotation. Alternatively, the membrane was incubated with 1 µg/mL of 6xHis Tag mouse antibody (HIS.H8, Thermo Fisher Scientific) and with 1:2000 of goat anti-Mouse IgG (H+L) secondary antibody (Thermo Fisher Scientific), for 1 hour at room temperature. For signal development, the membrane was covered with 1:1 of Clarity™ Western ECL blotting substrates (peroxide solution: luminol/enhancer solution) (Bio-Rad) and signal detection was carried out by using the ChemiDoc™ imaging system (Bio-Rad).

2.4.2. Large-scale production of hBMX in stirred-tank bioreactors

The hBMX large-scale production (Figure 2.2) was performed in a 5 L stirred-tank bioreactor system BIOSTAT® DCU-3 (Sartorius Stedim Biotech S.A., Germany). The dissolved oxygen (DO) concentration was controlled at 30% of air saturation by automatically varying the stirring rate between 60 and 130 rpm, and temperature was controlled at 27ºC. The inoculum was prepared by seeding 2 L shake flasks at 1x10^6 cells/mL in 250 mL of working volume. Cells were seeded into the bioreactor at a final concentration of 0.6x10^6 cells/mL and allowed to grow up to 1.0x 10^6 cell/mL. Cells were then infected using a MOI of 0.01 pfu/cell and harvesting was carried out at 83% of cell viability, corresponding to 72 hpi.

Figure 2.2. Large-scale production of hBMX in a 5 L stirred-tank bioreactor.

2.4.3. Downstream processing

2.4.3.1. Sample preparation

The infected cells were harvested after 72 hours post-infection by centrifugation at 500 g for 15 minutes at 4ºC. Pellets were re-suspended in five volumes of cold lysis buffer, consisting of 50 mM Hepes pH 8.0, 300 mM NaCl, 20 mM imidazole, 2 mM DTT, 1.25 ug/mL leupeptin, one tablet of complete EDTA-free Protease Inhibitor Cocktail per 100 mL of solution, 4U/mL of benzonase and 1 mM of $MgCl_2$. Cells were disrupted in an Avestin EmulsiFlex-C5 high-pressure homogenizer (Avestin, Ottawa, Canada), by two consecutive passes at an operating pressure of 500 MPa, each cycle lasting 5 minutes. The lysate was clarified by centrifugation at 30 000 g during 30 minutes at 4ºC and filtered with a 0.22 µm Millipore membrane.

2.4.3.2. Purification

The purification of the 6xHis-hBMX protein (32.5 kDa) was carried out following the guidelines described in ref. 15. The cleared lysate was loaded onto a 5 mL HisTrap™ HP column (GE Healthcare, Uppsala, Sweden), previously equilibrated with Buffer A (50 mM Hepes pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM DTT). The column was washed with buffer A and the immobilized his-tagged protein was eluted stepwise with Buffer B (50 mM Hepes pH 8.0, 300 mM NaCl, 1000 mM imidazole, 1 mM DTT). Selected peak fractions were injected onto a HiLoad™ 26/600 Superdex™ 200 column (GE Healthcare) equilibrated with 50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 1 mM DTT. The pooled fractions containing 6xHis-TEV-hBMX were treated with TEV protease (200 U/mg of protein) and 0.5 mM of EDTA was added. The mixture was incubated at 30ºC for 1h, followed overnight incubation at room

temperature. After the cleavage, the TVMV-treated sample was loaded onto a HiPrep™ 26/10 Desalting column (GE Heathcare) run with 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT to remove the traces of EDTA. The solution, containing both tagged and untagged species, was injected into a second 5 mL HisTrap™ HP column, equilibrated with Buffer A (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM imidazole, and 1mM of DTT) and protein fractions were eluted after 6 washing steps of Buffer B (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1000 mM imidazole, and 1mM of DTT). Peak fractions corresponding to both His_6-TEV-hBMX and hBMX were collected separately and applied onto a HiPrep™ 26/10 desalting column for buffer exchange to the crystallization buffer (20 mM Tris-HCl pH 8.0, 250 mM NaCl, 5 mM DTT). Final protein samples of His_6-TEV-hBMX and hBMX were stored at -80ºC. All experiments were run at room temperature and monitored by SDS-PAGE gel and Western Blot. SDS-PAGE was carried out using 10% Bis-Tris NuPage Gel (Invitrogen) and protein bands were visualized using InstantBlue (Expedeon Ltd., California, USA). For the Western Blot, bands were transferred to a PVDF membrane with iBlot2 system (Invitrogen). The membrane was rinsed in blocking buffer (3% skim milk from Sigma in 0.05% PBS Tween) and left at room temperature for 45 min in soft rotation. Alternatively, the membrane was incubated with 1:6000 dilution Mouse anti-His, (Sigma-Aldrich-H1029) and with 1:10000 dilution of Goat anti-mouse IgG Alkaline phosphatase (Sigma-Aldrich-A2429), each one for 1 hour at room temperature. For signal detection, the membrane was covered with SIGMAFAST BCIP/NBT Alkaline phosphatase substrate, one tablet in 20 mL demonized water.

2.4.4. Purified sample quality assessment by DLS and CD

2.4.4.1. Dynamic light scattering (DLS)

DLS for characterization of particle size measurements and detection of aggregates in purified samples was performed on a Malvern Zetasizer Nano ZS (Malvern Panalytical, Malvern, United Kingdom). Samples were diluted to a final concentration of 0.16 mg/mL in 1 mL of total volume and filtered with a 0.22 µm Millipore membrane. 400 µL of sample were transferred to a disposable solvent-resistant micro cuvette (ZEN0040, Malvern Store). Measurements were set to run for 2 hours and were carried out at 4ºC. Data collected by the DLS software was provided in terms of intensity, volume, and number distribution graphs along with statistical analysis for each.

2.4.4.2. Circular dichroism (CD)

The CD spectra of the protein were acquired using a J-815 CD spectrometer (Jasco, Easton, USA) in a 195-300 nm wavelength range. Protein buffer exchange to PBS was carried out prior to the CD experiment, using a PD-10 column. Samples were diluted to 0.5 and 0.25 mg/mL, and 200 µL were transferred to absorption cells of 1 mm path length (Hellma Analytics, Mülheim, Germany). Experiments were made at 25ºC and the final spectrum was an average of three consecutive scans, subtracted by a blank spectrum. Collected data was normalized by adjusting the values of ellipticity (θ) to molar concentrations.

2.5. Protein-ligand complex characterization

2.5.1. Differential scanning fluorimetry (DSF)

DSF experiments were carried out to measure the change in protein stability upon binding with inhibitor compounds. The assay was performed in MicroAmp™ EnduraPlate™ Optical 96-Well Clear Reaction Plates with Barcode (Applied Biosystems, Life Technologies, California, USA) using a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems). Pre-incubation of the protein with the compound for 2 hours at 4ºC was required prior to DSF experiments. The final reaction mixture (20 µL of total volume) contained 4 µg of 6xHis-hBMX, 4-fold of Protein Thermal Shift™ Dye (Applied Biosystems) diluted in protein buffer solution, and 100 µM of compound. The temperature was increased from 25ºC to 90ºC with an increment rate of 0.016ºC/s. Excitation and emissions filters were applied for Protein Thermal Shift™ Dye (470nm and 520nm, respectively) and for ROX reference dye (580nm and 623nm, respectively).The melting temperatures were obtained by taking the midpoint of each transition.

2.5.2. Surface Plasmon resonance (SPR)

SPR biosensor analysis was conducted on a Biacore 4000 (Biacore AB, GE Healthcare Life Sciences, Uppsala, Sweden). The 6xHis-hBMX catalytic domain construct was immobilized onto CM5 Series S sensor chip using covalent amine coupling. The carboxymethyldextran surface of all flow cells was activated with 20 mM EDC and 5 mM NHS for 2 min. HBS-N, which consisted of 10 mM Hepes, pH 7.4, 150 mM NaCl, was used as the immobilization buffer. Protonation of 6xHis-hBMX is required for protein immobilization, so that the surface and the protein carry opposite net charges. This was achieved by diluting the protein sample in 10 mM of sodium acetate, pH 5.5, to a final concentration of 10 µg/mL, and in the presence of 5µM staurosporine, a known kinase inhibitor[77]. The protein was coupled to the surface with a 2 min injection at a flow rate of 10 µL/min to immobilization levels of ca. 1500 and 3500 RU. The 6xHis-hBMX was directly dissolved in running buffer consisting of 20 mM Hepes, pH 7.4, 130 mM NaCl, 0.1 mM EGTA, 5 mM MgCl2, 1 mM DTT, 0.05% tween-20, and 2% of DMSO. The

inhibitor compounds were tested at 5 different concentrations using a 2 or 3-fold dilution series, being the highest concentration tested 1 µM or 0.5 µM (values optimized for each compound). Sample injection was carried out at 30 µl/min. All sensorgrams were processed by subtracting the binding response recorded from the buffer blank injection. All datasets were fit to a 1:1 interaction model to determine kinetic rate constants.

2.6. Protein-ligand complex determination by X-ray crystallography

2.6.1. hBMX co-crystallization with a covalent binder

High-throughput screens were set up to test a variety of commercial crystallization screens, including the Structure 1 and 2 (Molecular dimensions Ltd., Suffolk, UK), JCSG+™ (Molecular dimensions Ltd.), Morpheus® (Molecular dimensions Ltd.), Stura Footprint (Molecular dimensions Ltd.), BCS (Molecular dimensions Ltd.), and Salt Rx (Hampton Research, California, USA). The trials were carried out using the sitting-drop vapor diffusion method on the mosquito® LCP crystallization robot (TP Labtech Ltd, Hertfordshire, UK). The purified hBMX kinase was used for crystallization at a final concentration of 10 mg/mL (according to ref. 15) and pre-incubated for 2 hours at 20ºC with a 2-fold concentration of a covalent compound (code name: BMX-CI-651). The drops consisted of 0.150 µL of the reservoir solution mixed with an equal volume of the protein sample, and equilibrated against a 45 µL reservoir. Plates were incubated at 20ºC. Promising conditions that favored crystal formation were used for drop scale-up and the better-shaped crystals were analyzed at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France.

2.6.2. Model building and refinement

An X-ray diffraction data set to 2.0 Å was collected at ESRF beamline ID30A-3 with a Dectris EIGER X 4M detector from a cryocooled crystal at 100 K. The diffraction data were processed with AutoPROC and XDS[77], and the data processing and refinement statistics are summarized in Table 2.2. Two diffraction datasets were obtained: in the first, a spherical region of reciprocal space to 2.2 Å resolution was defined, and in the second a triaxial ellipsoidal region to a maximal resolution of 1.95 Å was selected with the STARANISO module of AutoPROC[77]. The structure of hBMX in complex with BMX-CI-651 was determined by molecular replacement with PHASER[78] as implemented in the CCP4 program suite[79,80] using the PDB entry 3SXS[15] as a search model, without including ligands and water molecules. Two independent copies of the search model were located in the crystal structure, and model rebuilding was carried out with BUCANEER[81] and COOT[82]. Initial structure refinement was undertaken with REFMAC[83].

Table 2.2. Data processing statistics for the crystallographic structure of hBMX in complex with BMX-CI-651.

Wavelength (Å)	0.9677
Temperature (K)	100
Scan type	ω
Total no. of frames	900
Width (°)	0.25
Total angular range (°)	225
Exposure time per image (s)	0.005
Space group	$P2_1$
Unit cell parameters	
a, b, c (Å)	66.7, 63.29, 74.96
β (°)	104.55
CCP4 TRUNCATE processing	
Resolution (°) [a]	63.3-2.19 (2.22-2.19)
Nr of observations	133203 (4732)
Unique reflections	30574 (1441)
Multiplicity	4.4 (3.3)
Completeness (%)	97.1 (91.7)
R-merge (%) [b]	23.6 (165)
R-p.i.m. (%) [c]	12.9 (108)
<I/σ (I)>	4.8 (0.7)
CC½	0.990 (0.394)
autoPROC STARANISO processing	
Resolution limits of ellipsoid fitted to resolution cut-off surface (Å):	
	1.98, 2.80, 2.25
Resolution, spherical limits (°) [a]	72.6-1.98 (2.24-1.98)
Nr of observations	99785 (3815)
Unique reflections	23531 (1161)
Multiplicity	4.2 (3.3)
Completeness, spherical (%)	56.0 (9.1)
Completeness, ellipsoidal (%)	87.6 (51.7)
R-merge (%) [b]	18.1 (76.0)
R-p.i.m. (%) [c]	10.1 (48.9)
<I/σ (I)>	6.1 (1.5)
CC½	0.991 (0.586)
Wilson B (Å²)	23.6
Z[d]	2
Estimated V_M [e]	2.46
Estimated solvent content (%) [e]	50.1

[a] Values in parentheses refer to the highest resolution shell; [b] R-merge=merging R-factor, $(\Sigma_{hkl} \Sigma_i I_i(hkl) - \langle I(hkl)\rangle I) / (\Sigma_{hkl} \Sigma_i I(hkl) \times 100\ \%$; [c] R-p.i.m. = precision-independent R-factor, $\Sigma_{hkl} [1/(N-1)]^{1/2} \Sigma_i I_i(hkl) - \langle I(hkl)\rangle I / (\Sigma_{hkl} \Sigma_i I_i(hkl) \times 100\ \%$. For each unique Bragg reflection with indices (hkl), I_i is the i-th observation of its intensity and N its multiplicity; [d] Nr. of molecules in the asymmetric unit; [e] According to Matthews coefficient.

The BMX-CI-651 ligand dictionary was created with JLIGAND[84] and the ligand was manually fitted into the electron density using COOT. Refinement was continued with PHENIX[85], alternating with manual model editing in COOT between refinements against σ_A-weighted $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ electron density maps. In the final refinement cycles, Translation-Libration-Screw rigid-body anisotropic atomic displacement parameters were refined, water molecules added automatically and the relative weights between the crystallographic and stereochemical energy terms optimized. Each hBMX molecule was divided into 4 rigid-body segments, estimated from the TLSMD server[86] using the isotropic atomic displacement parameters from a previous refinement run. The final refinement was carried out to 2.0 Å against the STARANISO dataset and the statistics are included in Table 2.3. Figures were prepared with PYMOL[87].

Table 2.3. Refinement statistics for the crystallographic structure of hBMX in complex with BMX-CI-65.

Resolution limits (Å) [a]	72.6 – 2.00 (2.09 – 2.00)
R_{work} [b]	23.2 (32.9)
R_{free} [c]	25.9 (44.6)
ML coordinate error estimate (Å) [d]	0.26
Model composition	
Non-hydrogen protein atoms	4308
Non-hydrogen ligand atoms	76
Solvent molecules	126
Mean B values (Å^2) [e]	
Protein main-chain	32.4
Protein side-chain	35.5
Ligands	25.4
Solvent	23.2
Model r.m.s. deviations from ideality	
Bond lengths (Å)	0.002
Bond angles (º)	0.603
Chiral centres (Å^3)	0.041
Planar groups (Å)	0.003
Model validation [i]	
% Ramachandran outliers	0.0
% Ramachandran favored	98.3
% Rotamer outliers	0.63
C^β outliers	0
Clash score	2.7

[a] Values in parentheses refer to the highest resolution shell; [b] $R_{work} = (\Sigma_{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)||) / (\Sigma_{hkl} |F_{obs}(hkl)|) \times 100$ %; [c] R_{free} is calculated as above from a random sample containing 5% of the total number of independent reflections measured; [d] Maximum-likelihood estimate by PHENIX; [e] Calculated from isotropic or equivalent isotropic B-values; [i] Calculated with MolProbity[88].

Chapter 3

3. Results and discussion

3.1. Analysis of the transformed competent *E. coli* cells by PCR

The Bac-to-Bac® Expression System was applied to the generation of a recombinant baculovirus which was later used for transfection and subsequent expression of the human recombinant hBMX in Sf-9 cells.

To generate the recombinant baculoviruses, competent *E. coli* cells were transformed with a donor plasmid containing the hBMX gene. The bacterial colonies that in fact harbored the hBMX gene appeared white in a background of blue colonies, due to the disruption of the LacZα gene. Figure 3.1 illustrates two inoculated plates with different bacterial cell densities. Overall, the transposition success rate of the hBMX gene in *E. coli* was very low. The values ranged from 20% in plate A (1 white/5 blue colonies) and 3% in plate B (3 white/106 blue colonies). The plate with less cell density (plate A) contained an isolated large white colony, whereas the plate with a higher cell density (plate B) had a large number of small colonies, in which the white ones could not be easily picked from the Petri dish.

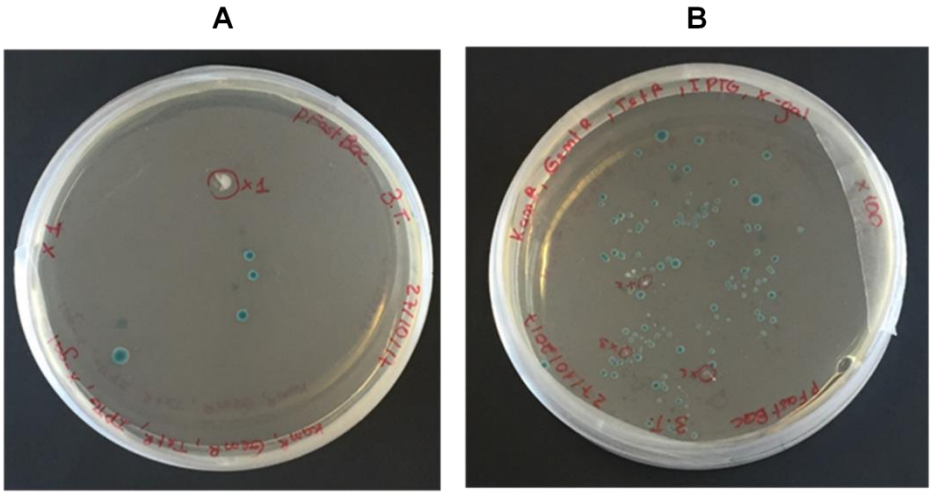

Figure 3.1. Plated DH10Bac™ *E. coli* cells after transformation proceedings. A) Transformation content concentrated to 1x; B) transformation content concentrated to 100x.

The recombinant bacmid DNA from the selected white colony (Figure 3.1, plate A) was purified with a final concentration of 1.2 µg/µL. A negative control was also performed by isolating the bacmid DNA from a blue colony. Confirmation of correct hBMX gene transposition was carried out by conventional PCR, resulting in the amplification of a single band around 1311bp (Figure 3.2). As expected, the bacmid DNA from a blue colony (negative control) was

not amplified. The absence of a band on the gel for the blue colony indicates that no recombination event occurred and therefore the hBMX gene is not present.

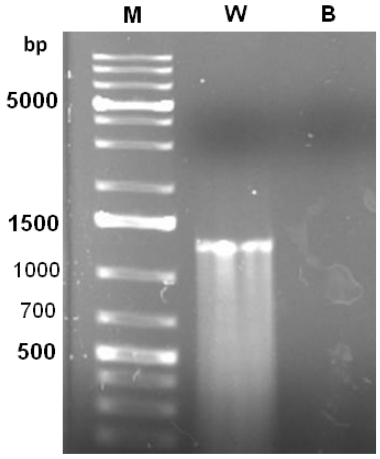

Figure 3.2. Analysis of the recombinant bacmid DNA. The DNA templates from both white (W) and blue (B) colonies used in the PCR reaction, were run on a 1% agarose gel in parallel to GeneRuler™ 1 kb Plus DNA marker (M). The PCR reaction of the recombinant bacmid DNA from the white colony resulted in amplification of a fragment with approximately 1311 bp – indicating an appropriate transposition of the hBMX gene. On the other hand, the bacmid DNA from the blue (negative control) was not amplified.

3.2. Insect cell transfection and protein production

Recombinant protein production using the baculovirus-insect cell expression system presents some advantages such as a very high production level of a protein, and mammalian-like post-translational processing and modifications.[89] However, the production of any protein via IC-BEVS requires a careful optimization of the process parameters to ensure batch consistency and product quality.[46,89] Therefore, pre-determining the optimum value for the multiplicity of infection (MOI) and for the time of harvesting (TOH), is extremely helpful to maximize productivity and product yield in a small or large-scale production of recombinant proteins using the Sf-9 cell line.[33,46,89]

As a prior step, the selected recombinant bacmid DNA containing the hBMX gene was used to transfect Sf-9 cells to produce recombinant baculoviruses. Results from microscopic observations using an invert phase microscope (Leica Microsystems, Wetzlar, Germany) revealed some characteristic phenotypical changes in transfected Sf-9 cells, such as ceasing of cell growth and decreasing of the cell viability (Figure 3.3). Increase in cell diameter and nuclei size, due to accumulation of viral replication products or altered cell organelles, was also observed (more clearly) around 72 hours post-transfection (hpt), followed by the loss of cell adherence to the culture plate. This type of cytophatic effect is commonly seen with adenoviruses, where infected cells become so significantly large that eventually detach from the surface.[89,90]

Once the cells appeared to be infected by showing all the characteristic signs of a late stage infection (usually around 72 hpt), the baculoviruses were harvested and used for viral amplification and titration.

The viral stock was amplified twice to generate higher titers for subsequent expression experiments. The titers of the P3 generation baculoviruses were calculated by the CGCA and by the MTT methods, and the results are presented in Table 3.1.

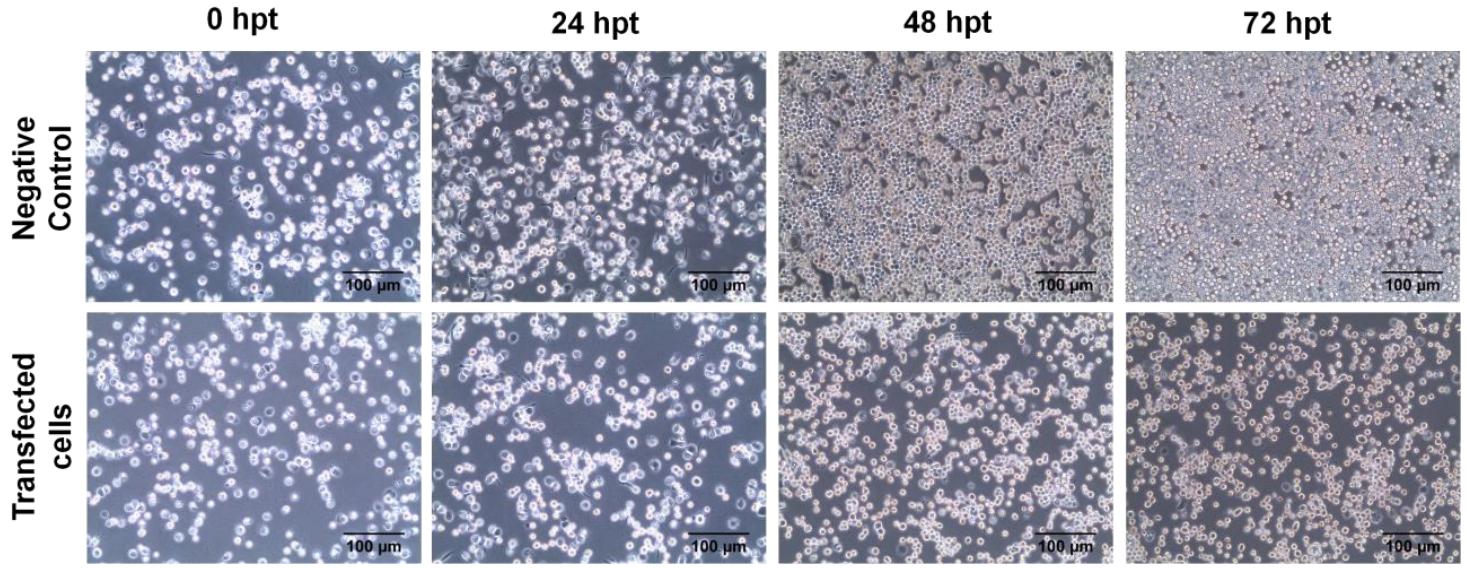

Figure 3.3. Microscopic view of Sf-9 cells transfected with the recombinant bacmid DNA. Cell shape, size and density was monitored immediately right after transfection (0 hpt) and for several hours post-transfection (24, 48, and 72 hpt). A negative control consisting of non-infected cells was also performed. Note the increase in cell concentration and nuclei size at 72 hpt, as well as the difference in cell density between the transfected and non-transfected cells.

Table 3.1. Baculovirus titers obtained with CGCA and MTT methods

Stock ID	CGCA (x10^7)	MTT (x10^7)
Bac#1.1 (P3)	1.58 ± 0.24	2.83 ± 0.29
Bac#2.1 (P3)	3.74 ± 0.56	2.03 ± 0.21

In both amplification experiments, we obtained titers of 10^7, which were in agreement to what is expected by the Bac-to-Bac® Baculovirus Expression System manual (ref. 43) : from $1x10^6$ to $1x10^7$ pfu/mL for the first amplification, and from $1x10^7$ to $1x10^8$ pfu/mL for the second amplification.

Next, a small scale feasibility test was performed and some of the cytophatic effects caused by the infection of the baculovirus in Sf-9 cells were registered (Figure 3.4). Examined parameters, MOI and TOH, were explored to determine the optimum MOI value (1, 0.1, 0.01, or 0.001 pfu/cell) and harvest time (0, 24, 48, 72, or 96 hpi) for maximum protein production and yield. Cell cultures were also monitored daily based on viable cell concentration, percentage of viability and cell size, until 96 hpi. Overall, the results were in agreement with prior microscopic observations. Both viable cell concentration and the percentage of viability decreased significantly (as clearly seen at 72 hpi), and cells suffered an increment of ca. 2-3 µm in size. In contrast, non-infected cells continued to grow exponentially after 72 hours and no significant variation in cell size was observed. Furthermore, MOI values showed to directly influence cell growth in an inverse way. At high MOIs, cells suffered a quicker growth cessation, resulting in a lower maximum of viable cell density; whereas in low MOIs, a high maximum of viable cell density was reached and the onset of cell death was delayed. Cessation of growth observed in higher MOIs occurs due to the almost complete infection of the entire cell population. In low MOIs, however, only a small proportion of cells are actually infected, and therefore only these cells stop dividing; meanwhile the non-infected cells continue to grow at a normal rate after the infection.[91,92] Subsequently, in low MOIs, cell growth is eventually arrested, due to the secondary absorption and infection of non-infected cell population by the viral particles that had successfully replicated in the first round of infection.[92]

To examine the quantity of intracellular hBMX produced by the baculovirus-infected cells, we analyzed the samples that were taken daily in the small scale feasibility test by SDS-PAGE gel and Western Blot (Figure 3.5). In agreement with previous results obtained and the literature (refs.15, 43, 46, 77, 79, and 80), two different MOI values were tested (0.01 pfu/cell and 0.001 pfu/cell), as well as two different time points of sampling (48 and 72 hpi). The intracellular hBMX protein was detected at both MOIs and time points; however, there is a slightly greater production of hBMX at 72 hpi, which correlates with the time of cell harvesting supported by the literature. The detection of the human recombinant hBMX is not possible on the first 18-24 hpi, since the expression of the foreign gene is driven by the *polh* promoter, which is only transcribed during the late stages of infection;[30,36] and after 72 hpi, the entire cell population reaches the decline phase (Figure 3.4, A and B), with consequent cell lysis and release of the recombinant protein, which may result in product contamination and degradation.[92] Moreover, the amount of recombinant product expressed is accumulative over time and appears to be the same for both MOI values at 72 hpi. This is could be explained by the similar percentage of post-infection growth experienced by non-infected cells at both MOIs, resulting in a similar amount of intracellular product expressed.

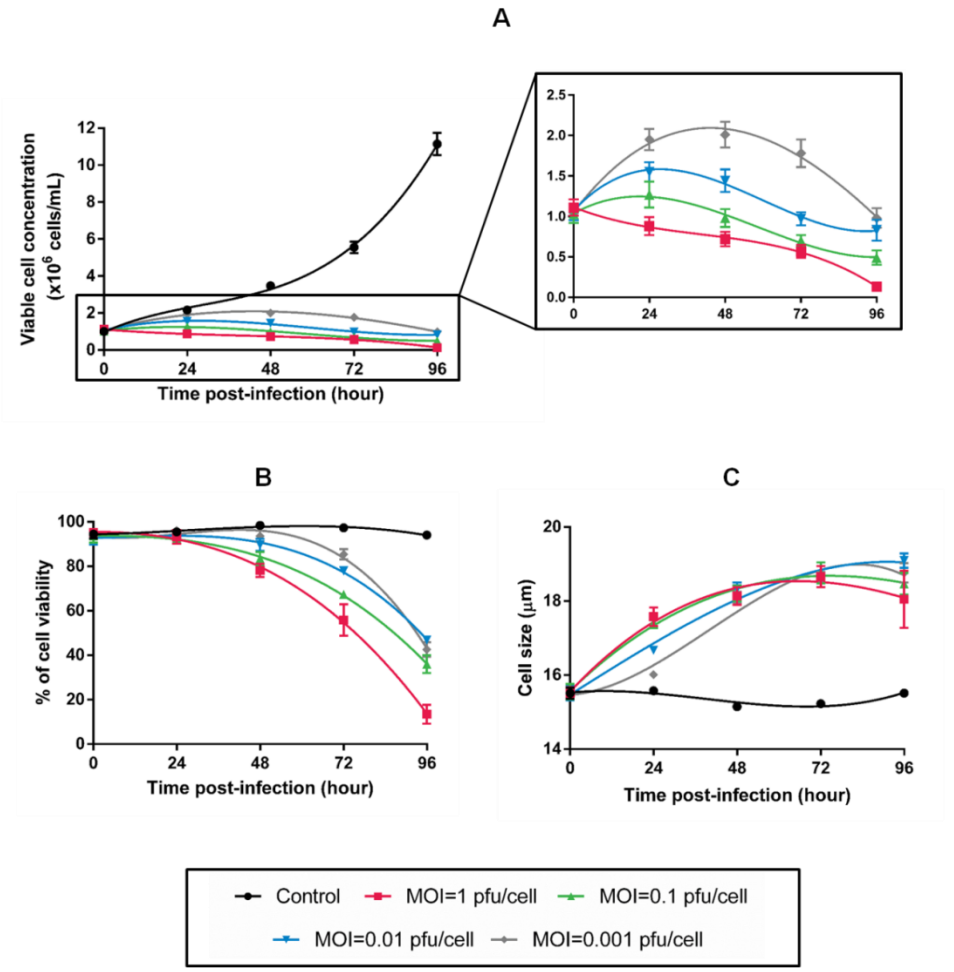

Figure 3.4. Graphic illustrations of the daily monitoring of some cytophatic effects caused by the baculovirus infection. A) Viable cell concentration (x10^6 cells/mL); B) Percentage of cell viability; C) Cell size (µm).

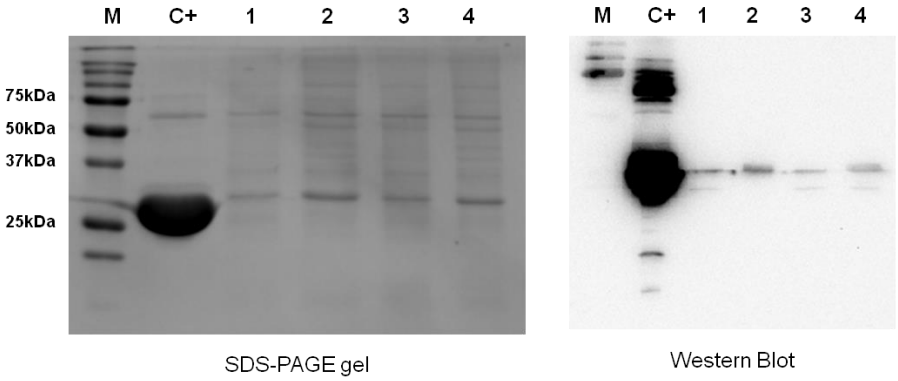

Figure 3.5. SDS-PAGE gel and Western Blot for hBMX detection. The SDS-PAGE and Western Blot were performed using the pellet of cells infected with a MOI of 0.01 pfu/cell (lanes 1 and 2) and of 0.001 pfu/cell (lanes 3 and 4). The presence of hBMX in cell lysate was confirmed at 48 hpi (lanes 1 and 3) and at 72 hpi (lanes 2 and 4). SDS-page was run on a 4-15% acrylamide gel in parallel with Precision Plus Protein™ All Blue Prestained Protein Standards (M). A positive control was also performed by loading a 6xHis-tag protein into the gel (C+). The Western Blot was carried out on the baculovirus infected Sf-9 cells with the 6x-His Tag mouse primary antibody and with the goat anti-Mouse IgG (H+L) secondary antibody.

Following the results obtained in the small scale feasibility test, we proceeded for the large-scale production of the human recombinant BMX in a 5 L stirred-tank bioreactor. Figure 3.6

illustrates the daily monitoring of the cells in the bioreactor based on viable cell concentration, percentage of viability and cell size, until 72 hpi. Reproducible results were obtained in the 5 L STB when compared with the small scale feasibility test, as well as in the infected-control culture shake flask. Overall, viable cell concentration and percentage of viability were decreased significantly and cell size increased ca. 2-3 µm, as previously reported in the small scale feasibility test.

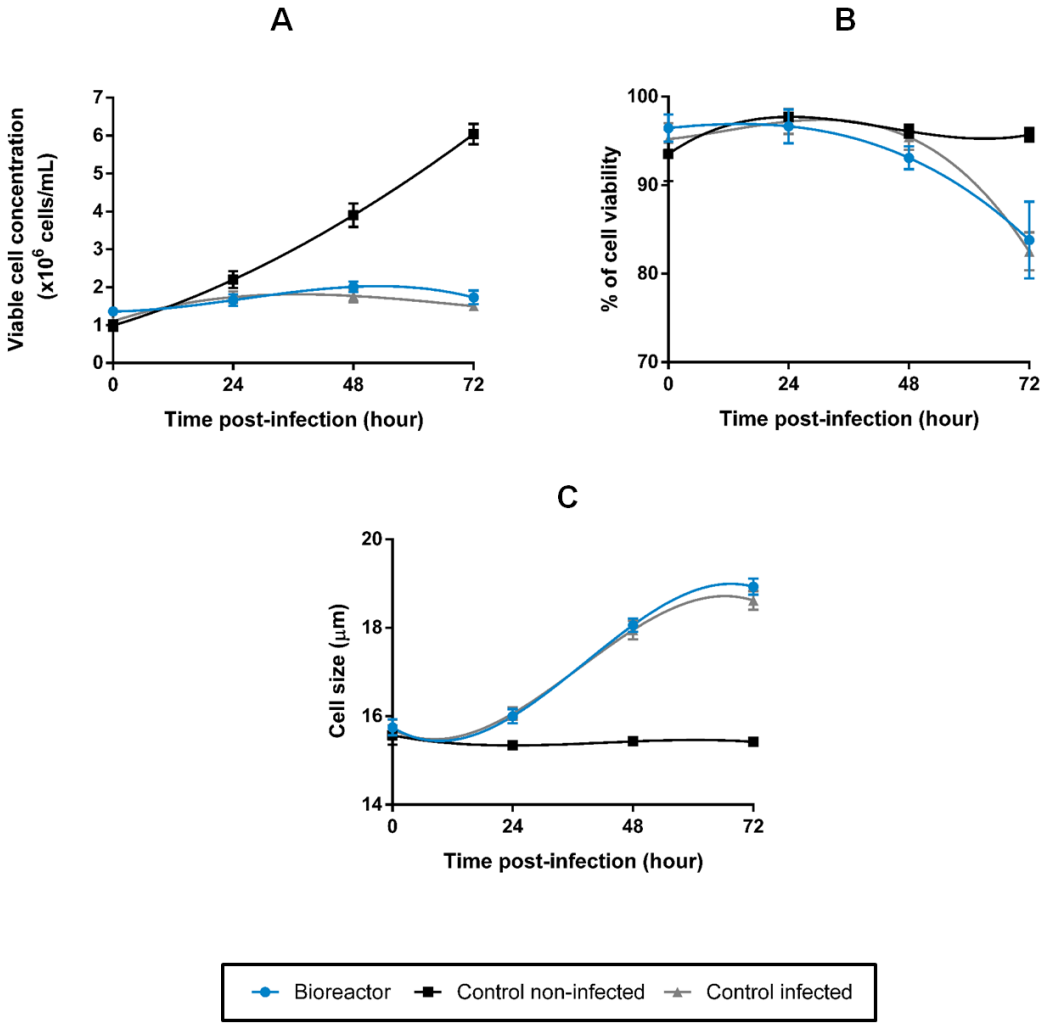

Figure 3.6. Graphic illustrations of the daily monitoring of the cells in the bioreactor. A) Viable cell concentration (x10^6 cells/mL); B) Percentage of cell viability; C) Cell size (µm). Positive (infected) and negative (non-infected) controls were also performed.

3.3. Purification process

The recombinant hBMX was expressed with a 6xHistine tag fused to its N-terminal for high affinity binding to nickel-charged columns. After the first purification step using a His-Trap affinity column, the fractions with hBMX protein sample were further purified using a Superdex 200 column (Figure 3.7, A) to remove aggregates and other residues and a single peak was detected. The fractions from the elution peak were analyzed by SDS-PAGE gel (Figure 3.7, B),

and treated, subsequently, with TEV protease, first for 30 minutes at 30ºC, and then overnight at 24ºC. After TEV digestion and buffer exchange to remove the traces of EDTA, the sample was loaded onto a His-Trap affinity column and eluted stepwise with increasing imidazole concentrations. Six peaks were detected, one of them corresponding to the untagged protein (Peak U, Figure 3.8, A), and the remaining corresponding to the 6xHis-tagged hBMX. The fractions were analyzed by Western Blot (Figure 3.8, B) and the fractions that displayed the highest purity were pooled and concentrated for subsequent experiments.

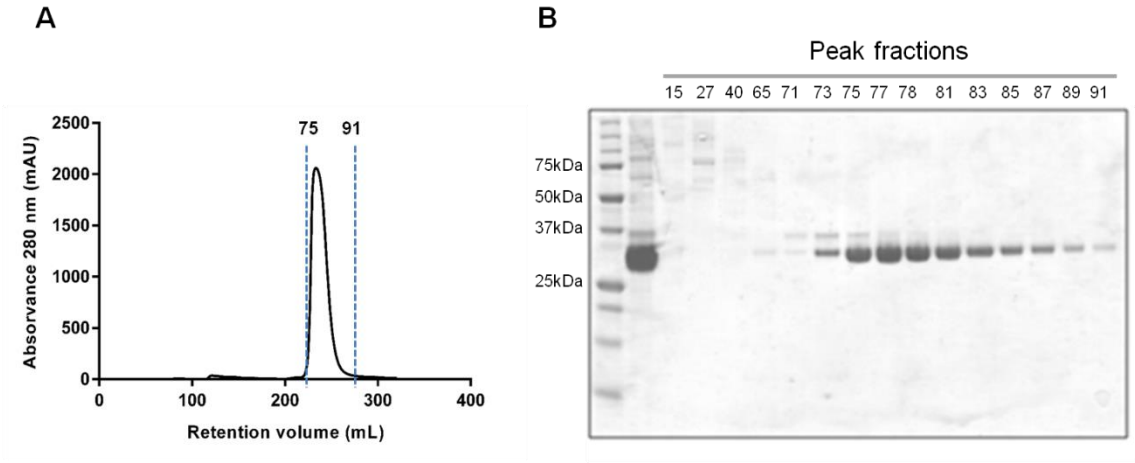

Figure 3.7. Size exclusion chromatography of the hBMX protein and SDS-PAGE gel analysis. A) Size exclusion chromatography. Only a single peak was observed. B) SDS-PAGE gel analysis. Lanes 15 to 91 denote the peak fractions obtained during the chromatographic experiment.

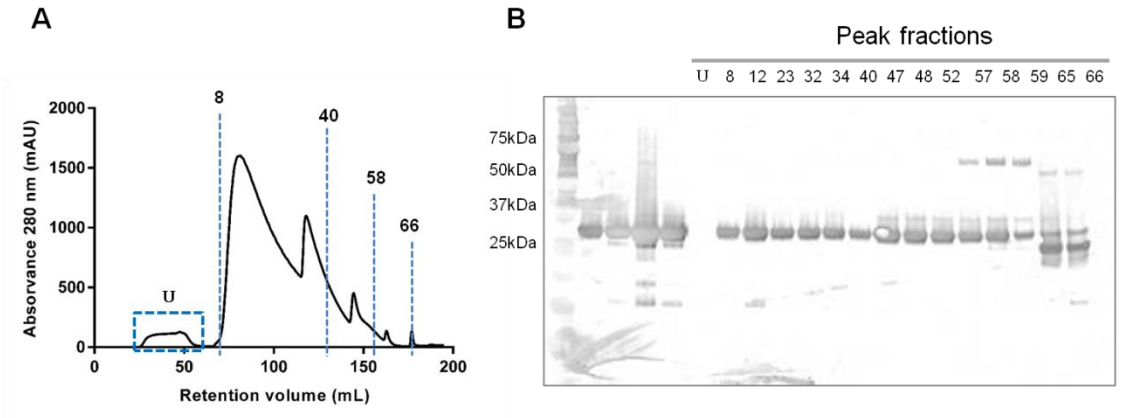

Figure 3.8. Affinity chromatography of the 6xHis-hBMX protein after TEV digestion and Western Blot analysis. A) Affinity chromatography. Protein was eluted with different percentages of imidazole, resulting in 6 peaks: peak U corresponds to the untagged hBMX; the remaining peaks correspond to the 6xHis-tagged protein. B) Western Blot analysis. Lanes U to 66 denote the peak fractions obtained during the chromatographic experiment. The untagged protein is not present in lane U, whereas the 6xHis-tagged protein is observed on the lanes 8 to 66.

Analysis of the final purification products, obtained by Western Blot, revealed some contamination in the pooled fractions of untagged hBMX and some degradation in the 6xHis-hBMX fractions (Figure 3.9). The aggregation problem was later solved by adding an ultracentrifugation step at 42 000 rpm, for 30 minutes, at 4ºC.

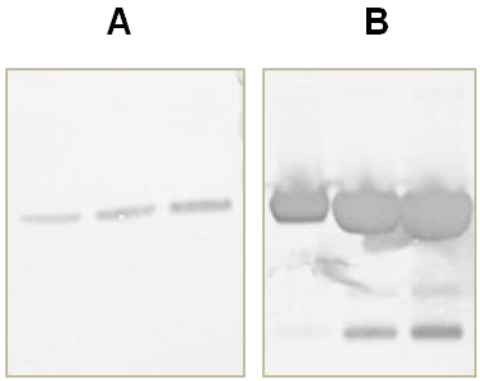

Figure 3.9. Analysis of the product obtained in the end of the purification process by Western Blot. A) Untagged hBMX protein. B) 6xHis-tagged hBMX protein. Note that the untagged protein is contaminated by the 6xHis-tagged protein and there is also some degradation in the 6xHis-tagged protein sample.

Table 3.2. Summary of purification parameters of the recombinant hBMX.

Chromatographic column	% Buffer B	Elution volume (mL)	Q_T (mg)/280 nm
1) HisTrap™ HP 5 mL	0, 8, and 20	37, 59, and 90	157.8
2) HiLoad 26/600 Superdex 200	N/A	234	111.67
3) HiPrep 26/10 Desalting	N/A	23	-
4) HisTrap™ HP 5 mL	0, 2, 3, 5, 10, and 50	48, 81, 118, 144, 163, and 178	7.74[a] and 76.98[b]
5) HiPrep 26/10 Desalting	N/A	37[a] and 39[b]	2.40[a] and 74.50[b]

[a] Untagged hBMX

[b] 6xHis-tagged hBMX

After several purification steps, we obtained a total of 2.4 mg of untagged hBMX and 74.5 mg of 6xHis-tagged hBMX, corresponding to a final yield of 15.4 mg of purified protein per each litter of cell culture. Table 3.2 summarizes the percentages of the Buffer B used, the elution volumes (mL), and the total quantity (Q_T; mg) of protein at the end of each purification step.

3.3.1. Purified sample quality assessment by DLS and CD

The reason why protein crystal formation is a major challenge is related with the large number of physiochemical and biochemical variables that impact the thermodynamics of this process. Homogeneity, chemical purity, stability, aggregation and structural state of a protein in solution, are some of the variables that may influence protein crystallization. Methods such as DLS or CD can provide useful information on these variables, and they can help to understand in which conditions a protein is more stable or starts to denature.[50,51]

The first DLS measurement was performed on a purified sample of 6xHis-hBMX to evaluate the dispersion level of the particles in solution and their aggregation state. The sample was previously filtered and the experiment was carried out at 4ºC for 2 hours. Throughout the experiment, the measured average size distribution of hBMX in solution was not consistent and the sample was aggregating overtime (Figure 3.10, A). Therefore, the sample is said to be polydisperse and unstable, and it is most likely not to crystallize.

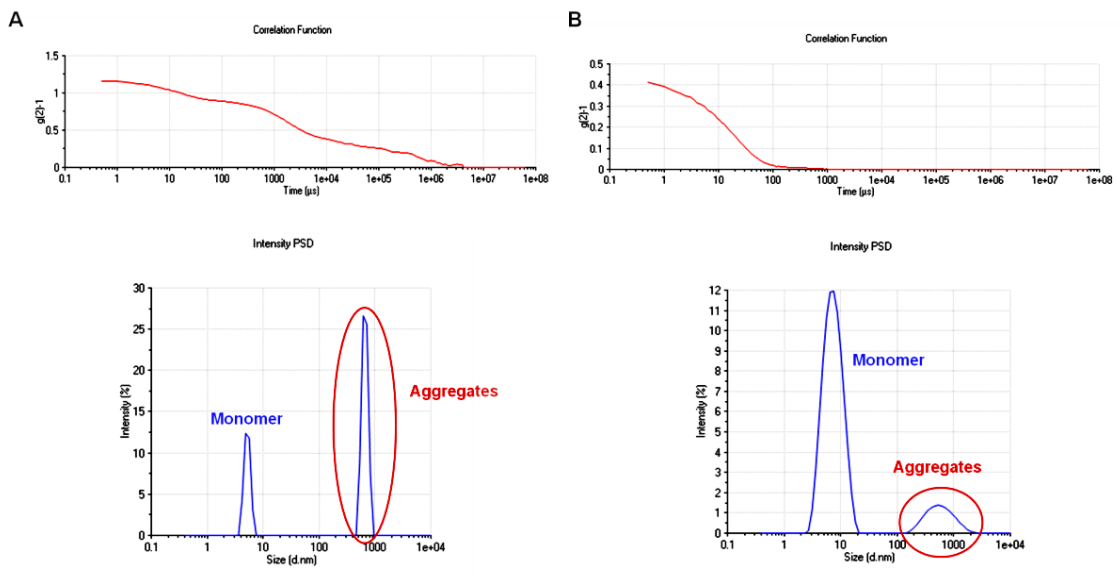

Figure 3.10. DLS autocorrelation function and spectrum of hBMX, where the relative intensity (%) is displayed as a function of the diameter of the protein particle. A) Analysis carried out at on a sample not ultracentrifuged. B) Analysis of the same sample after ultracentrifugation.

A second DLS measurement was carried out (Figure 3.10, B), but this time the protein sample was previously ultracentrifugated at 42 000 rpm, for 30 minutes at 4ºC. The correlation function displayed a typical curve for monodisperse samples and the size distribution of the

particles in solution was consistent with the size of the monomer, meaning that we were able to improve the quality of the sample. Despite these efforts, we did not manage to eliminate completely all the aggregates and other impurities that were interfering with the measurement.

The observed aggregation may be due to the mutagenesis strategy employed to the hBMX protein sequence (information mentioned in ref. 15). This strategy focused on replacing non-conserved residues by other residues that mimicked the crystal contacts of BTK, with the goal of creating packing opportunities for hBMX crystal formation. Therefore, DSF experiments were set up to optimize the buffer conditions referenced in the literature (ref. 15), as a way of improving protein stability in solution and minimize protein aggregation (data not shown). Overall, we could not find a better buffer condition or additive that could help stabilize the protein in solution, and therefore we proceeded for the crystallization process with the referenced buffer.

The conformational and folding states of 6xHis-hBMX were determined by CD in the far UV region (195-300 nm). The hBMX spectrum displays a fourth degree curve typical of proteins that have a predominance of α-helices in their composition (Figure 3.11). Indeed, hBMX kinase (published data; PDB entries: 3SXR and 3SXS) is an α-rich protein, containing ca. 86% of α-helixes in its structural composition. Furthermore, there are no signs of denatured sample, and therefore, the protein is in its native folded state.

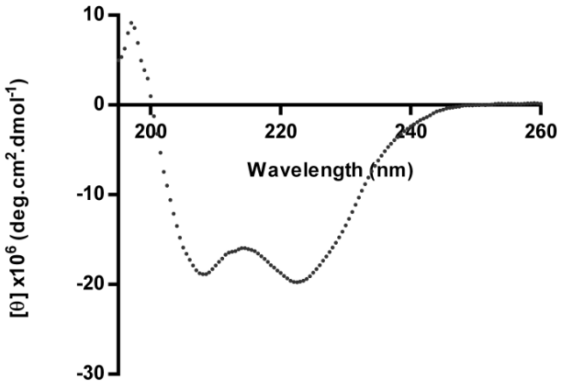

Figure 3.11. Characterization of hBMX by circular dichroism. The protein displays a curve typical of predominantly α-helical structures.

3.4. Protein-ligand complex characterization by DSF and SPR

Affinity-based biophysical methods are important tools for detection and analysis of the molecular interactions between proteins and small molecules. This enables the identification and characterization of improved drug candidates, which can desirably modulate the activity of a target protein for therapeutic applications.[94,95]

In the purpose of developing new hBMX inhibitors, particularly for the treatment of prostate cancer, our work focused on characterizing protein-ligand complex interactions using two different methodologies: DSF and SPR.

As a first round of selection, 25 different compounds with a similar chemical structure to BMX-IN-1 (mentioned in ref. 28) were analyzed by DSF, including the BMX-IN-1 for comparison. From this experiment, 3 major hits were obtained: BMX-CI-651, BMX-CI-51, and BMX-CI-61, which showed a greater capability to stabilize the active structure of hBMX than BMX-IN-1 (Figure 3.12). Indeed, while the melting temperature increase of hBMX with BMX-IN-1 was ca. 8ºC, a further increase of about 1 to 3ºC was observed in the presence of BMX-CI-651, BMX-CI-51 and BMX-CI-61.

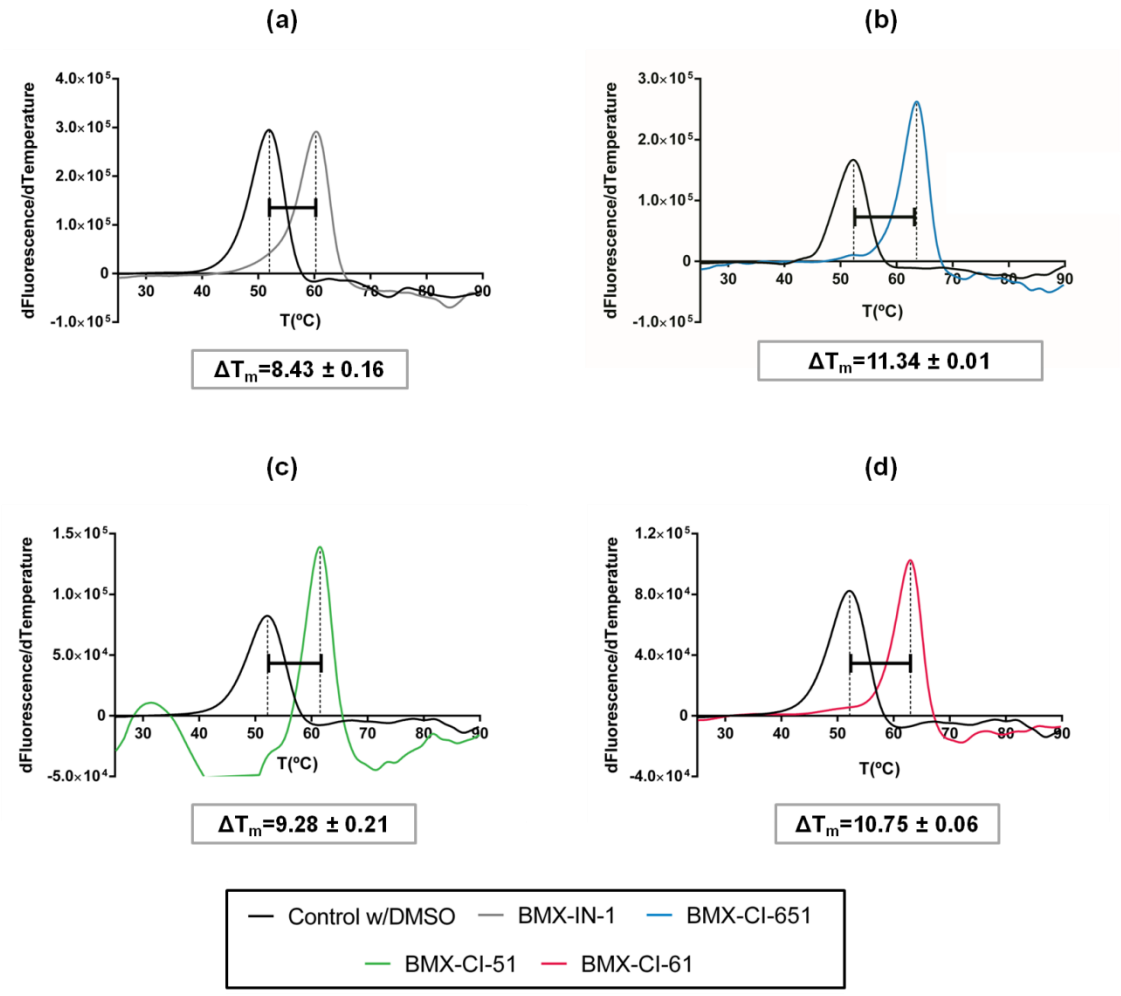

Figure 3.12. Thermal melting derivative curves of hBMX in the presence of 4 different covalent binders: (a) BMX-IN-1; (b) BMX-CI-651; (c) BMX-CI-51; (d) BMX-CI-61. The ΔT_m represents the shift in the melting temperature curve caused by the presence of a binder molecule.

For SPR studies, 6xHis-hBMX was immobilized on a CM5 chip at a final concentration of 10 µg/mL and directly dissolved in running buffer. The 4 covalent binders were tested at 5 consecutive dilutions of 2 or 3-fold series, being the highest concentrations 1 µM or 0.5 µM. The

sensorgrams obtained for the interactions of these compounds with hBMX, are shown in Figure 3.13.

Overall, the binding responses (RU) varied in a concentration-dependent manner; however, the plateau responses illustrated on the four sensorgrams do not depict a full binding cycle, since the response curves do not return to the baseline. These results are indicative of the irreversible nature of all compounds tested, thus remaining bound to the protein even after injection of the dissociation buffer for 1000s or 2000s.

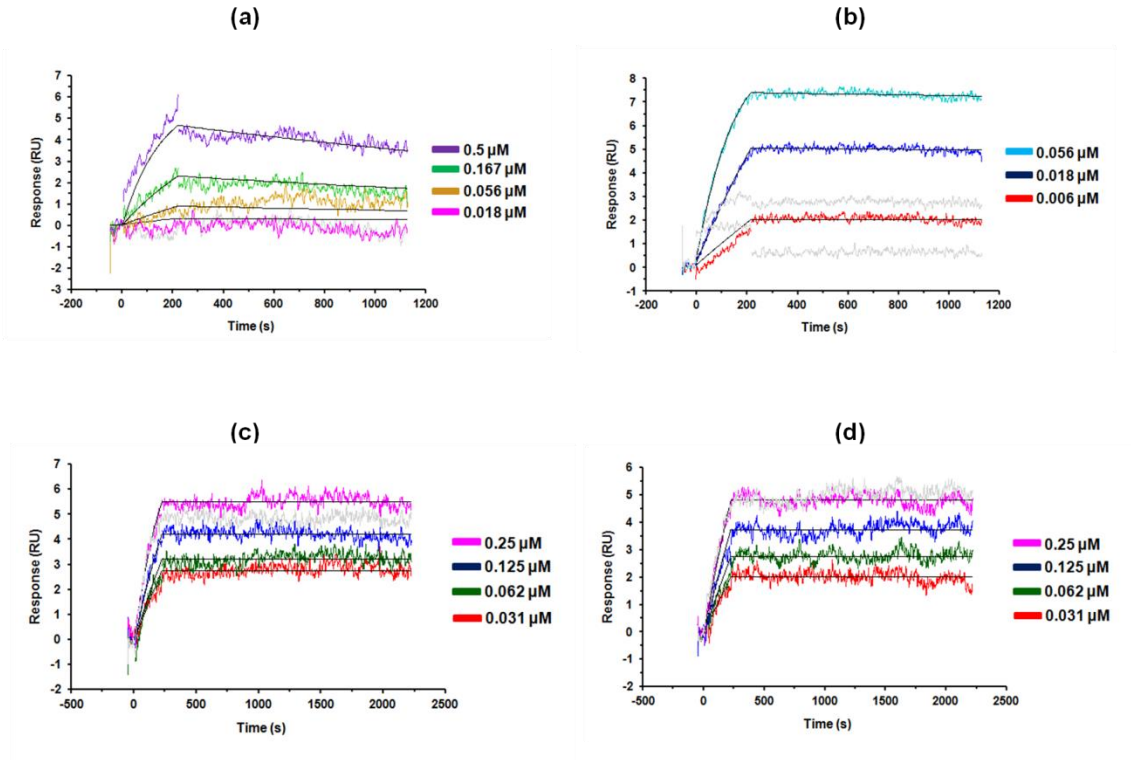

Figure 3.13. SPR binding analysis of hBMX catalytic domain with 4 different covalent compounds: (a) BMX-IN-1 (reference compound); (b) BMX-CI-651; (c) BMX-CI-51; (d) BMX-CI-61. Each graphic depicts the response upon the injection (t=0 seconds) of a compound at 5 different dilutions, and the response of subsequent washing step (t=200 seconds). The raw data of individual measurements are shown in color and the fitted data in black. Note that the compounds BMX-CI-651, BMX-CI-51, and BMX-CI-61, do not dissociate after injection of the dissociation buffer.

In order to evaluate the binding kinetics parameters for all 4 compounds, we determined the association and dissociation constants by fitting the raw data of individual measurements (Figure 3.13, colored lines) into a 1:1 interaction model (Figure 3.13, black lines). Calculated K_D values ranged the picomolar units for the newly designed compound, BMX-CI-651 (K_D = 143 pM), and the nanomolar range for BMX-IN-1 (K_D =69.7 nM), which is in accordance with the IC50 of 25 nM already published (ref. 28) (Table 3.3).

Table 3.3. SPR binding parameters for the interaction of hBMX with different covalent binders.

Compound	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	$k_D=k_{off}/k_{on}$ (nM)
BMX-IN-1	$7.35x10^{+3}$	$5.10x10^{-4}$	69.40
BMX-CI-651	$1.42x10^{+5}$	$2.03x10^{-5}$	0.143
BMX-CI-51	$7.16x10^{+4}$	< 0.0001	ND
BMX-CI-61	$9.94x10^{+4}$	< 0.0001	ND

ND - not determined

In addition to BMX-CI-651, both BMX-CI-51 and BMX-CI-61 have also showed a greater affinity towards BMX than BMX-IN-1. However, since their K_{off} values are really close to the limit of detection of the instrument, it is not possible to determine the real value.

3.5. Protein-ligand complex structure determination by X-ray crystallography

3.5.1. hBMX crystallization

To date, only two structures of the human BMX catalytic domain were deposited on PDB (3SRX and 3SXS)[15], both of them being hBMX complexes with known kinase inhibitors: Dasatinib and PP2. Although two referenced hit conditions were known to start with, the task of crystallizing this protein was very challenging.

As a first approach, we set up crystallization trials with 6xHis-hBMX alone or pre-incubated with BMX-CI-651 and used the conditions reported in Muckelbauer *et al.*, in 2011 (ref. 15). Other protein concentrations were also tested, such as 7.5, 12.5, and 15 mg/mL, as well as different pH values and concentrations of the main precipitants (ammonium sulfate and PEG 10000). From these first trials, the most observed outcomes were either precipitate or spherulites.

Next, it was decided to test a variety of commercial conditions by carrying out the experiments on a crystallization robot. This time, 1 hit condition was obtained, consisting of 0.2 M imidazole-malate buffer, pH 5.5, with 42% v/v PEG 600 (Figure 3.14). The crystals appeared after 2 days and reached their full size in 1 week. These crystals were cryoprotected with a solution containing 0.2 M imidazole-malate buffer, pH 5.5, with 42% v/v PEG 600 and 25% glycerol, flash-cooled in liquid nitrogen, and later sent to the ESRF for X-ray data collection.

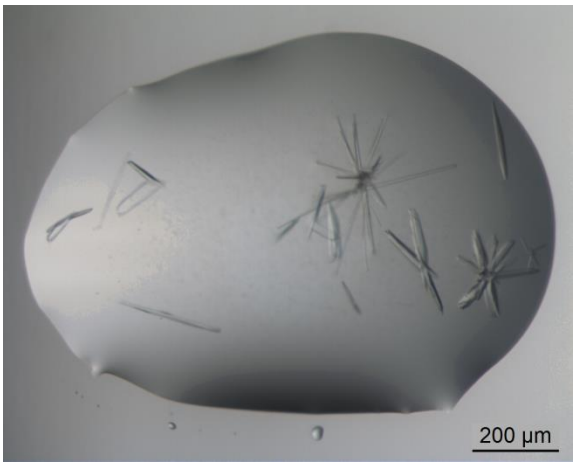

Figure 3.14. Optical microscopy image of hBMX crystals in complex with BMX-CI-651.

3.5.2. Overall structure of hBMX

The X-ray crystal structure of 6xHis-hBMX kinase catalytic domain in complex with BMX-CI-651 has been determined to 2.0 Å resolution (Figure 3.15). The crystal structure of hBMX/BMX-CI-651 contains two independent protein chains in the asymmetric unit, and each chain is covalently bound to one BMX-CI-651 ligand molecule. Chain A includes residues from 408 to 671 and chain B includes the residues from 411 to 671. Residues 408 to 410 in chain A belong to the uncleaved His-tag region.

As described in section 1.2.1, the hBMX tyrosine kinase catalytic domain folds into a typical kinase bi-lobed structure, with the ATP binding pocket situated in a cleft between the carboxy- and amino- lobes. The smaller amino- lobe (residues 411- 497) consists predominantly of β-strands, with one αC-helix (residues 454 - 466), and also contains a G-loop (residues 417-436) and a hinge region (residues 490 - 497). The carboxy- lobe (498 - 671) consists mostly of α-helices, a few short β-strands, and it also includes a DFG motif (residues 554 - 556) and an activation loop (residues 557 - 576).

The superposition of both hBMX/BMX-CI-651 chains and of these with BMX/Dasatinib (PDB 3SXR) and BMX/PP2 (PDB 3SXS) structures, was carried out with LSQKAB[96] in the CCP4 suite. The root mean square deviations (r.m.s.d.), which resulted from superposing the Cα's in the each pair of structures, are listed in Table 3.4. Overall, the superposition of the two hBMX/BMX-CI-651 molecules in the asymmetric unit (Figure 3.16, A), demonstrated that both chains have a very similar fold, except for some of the regions in the amino-lobe, particularly the αC-helix and the G-loop. The r.m.s.d. determined were higher when comparing the two amino-lobes separately (1.06 Å), than when comparing both carboxy-lobes (0.21 Å). Superposition of the hBMX/BMX-CI-651 chains with those of BMX/Dasatinib and BMX/PP2 also showed a nearly identical fold for the carboxyl-lobes, but with some major core differences in the amino-lobes (Table 3.4 and Figure 3.16, B).

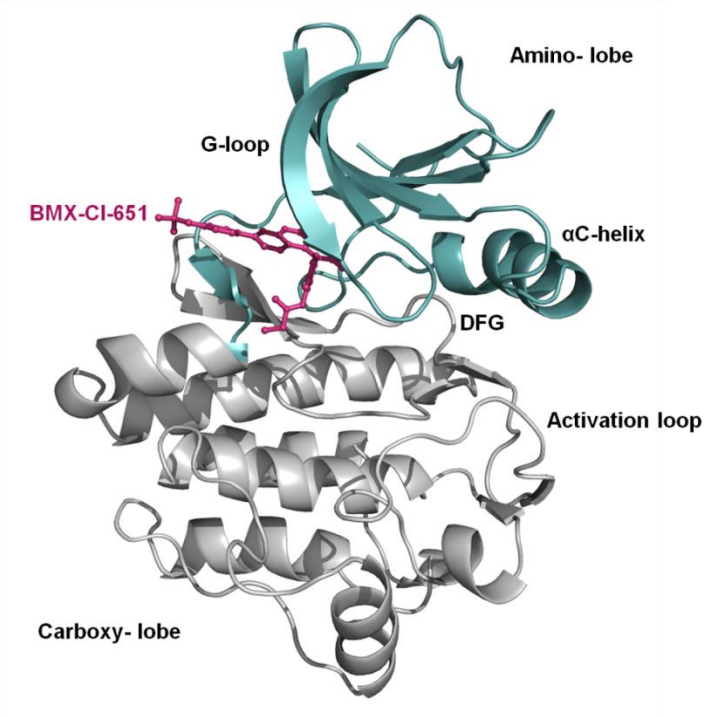

Figure 3.15. The overall structure of the hBMX kinase catalytic domain in complex with BMX-CI-651. The crystal structure of hBMX is represented as teal (for the amino-lobe) and grey (for the carboxy-lobe) cartoon. The covalent inhibitor is colored magenta and shown in ball and sticks representation.

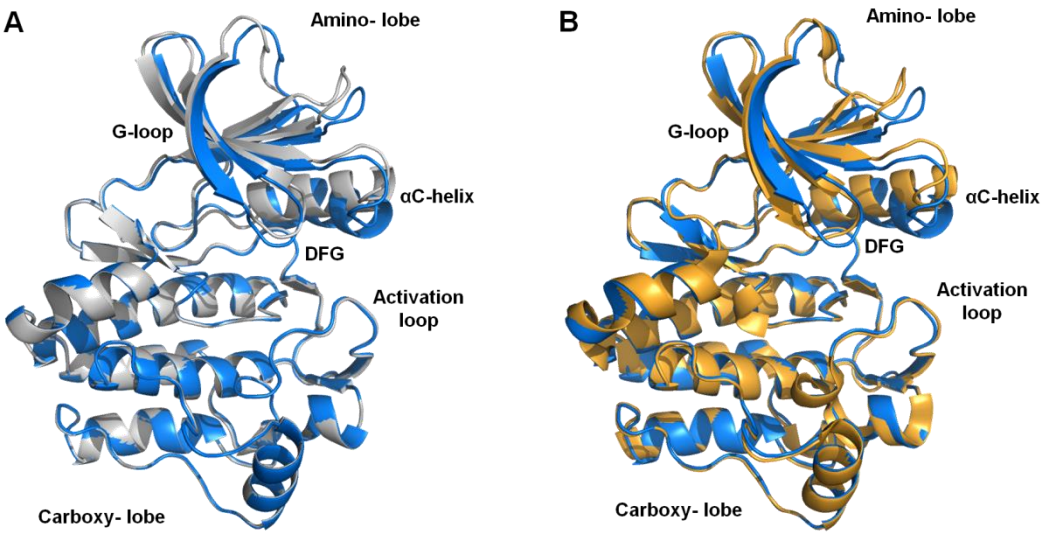

Figure 3.16. Alignment of the hBMX kinase catalytic domain structures. A) Superposition of the two independent chains of the hBMX/BMX-CI-651 complex structure. Chain A is colored in marine blue and chain B in grey. B) Superposition of hBMX/BMX-CI-651 complex with BMX/PP2. Chain A of hBMX/BMX-CI-651 is colored in marine blue and the BMX/PP2 complex is represented in orange.

The changes in BMX conformation are mainly caused by the existence of relatively flexible regions in the amino-lobe, specifically, the G-loop and the αC-helix. These conformational variations are also observed amongst the TEC and the Src family structures, meaning that the ATP binding site can adopt different conformations in order to accommodate different size and shaped ligands.[15] The apparent stable DFG-motif and the activation loop are also quite flexible regions in the BMX structure, but in this case they remain unchanged, revealing that the mechanism of BMX inhibition between the BMX-CI-651, Dasatinib, and PP2, has some similarities. Nevertheless, the core differences observed in the two hBMX/BMX-CI-651 molecules (chain A and B), suggest that the packing of the molecules within the crystal may have also influenced the conformational differences observed between the amino-lobes of the two independent chains.

Table 3.4. Superposition of hBMX/BMX-CI-651 with hBMX/Dasatinib and hBMX/PP2 complex structures.

		hBMX/BMX-CI-651	
		Chain A	**Chain B**
hBMX/BMX-CI-651	**Chain A**	N/A	**Global:** 0.84 Å **Amino-lobe:** 1.06 Å **Carboxy-lobe:**0.21 Å
	Chain B	-	N/A
hBMX/Dasatinib	**Chain A**	**Global:** 1.23 Å **Amino-lobe:** 1.03 Å **Carboxy-lobe:** 0.63 Å	**Global:** 0.85 Å **Amino-lobe:** - **Carboxy-lobe:** -
	Chain B	**Global:** 1.25 Å **Amino-lobe:** 1.03 Å **Carboxy-lobe:** 0.62 Å	**Global:** 0.84 Å **Amino-lobe:** - **Carboxy-lobe:** -
hBMX/PP2	**Single chain**	**Global:** 0.92 Å **Amino-lobe:** - **Carboxy-lobe:** -	**Global:** 0.58 Å **Amino-lobe:** - **Carboxy-lobe:** -

The carboxy-lobe regions of the hBMX/BMX-CI-651 structure were in line with the BMX/Dasatinib and BMX/PP2 complex structures. One of these regions included the activation loop which is in its outermost conformation, and it is similar to the activation loop of apo BTK protein.[15,97]

Despite the fact that the activation loop is at its outermost conformation, the structure of hBMX catalytic domain renders an inactive kinase conformation. This is mostly due to the DFG residues, specifically Asp554 and Phe555, whose conformation blocks the interaction of the

ATP-substrate with the binding determinants within the hBMX catalytic center.[15,98] In an active protein kinase condition, the DFG-motif adopts an in-like conformation, where the DFG Asp points inward to the ATP binding pocket and helps to position the γ-phosphate for its transfer.[15,98,99] The aromatic side chain of DFG Phe packs against the αC-helix and facilitates the formation of a salt bridge between a conserved lysine, located in the β3 strand, and a glutamate residue in the αC-helix (Figure 3.17, A).[15,98,100] In hBMX/BMX-CI-651, however, the DFG Asp554 side chain is positioned in the back cleft, away from the ATP binding pocket, and the DFG Phe555 aromatic ring points up into the gatekeeper region blocking the β3 Lys445-αC Glu460 ion pair formation (Figure 3.17, B). The DFG-out-like conformation found in hBMX/BMX-CI-651 positions the Phe and the Asp in a way similar to the DFG-out conformation observed in type II inhibitor complexes.[15,98] However, in hBMX the DFG-motif is relatively closed to αC-helix and the hydrophobic pocket is not present.

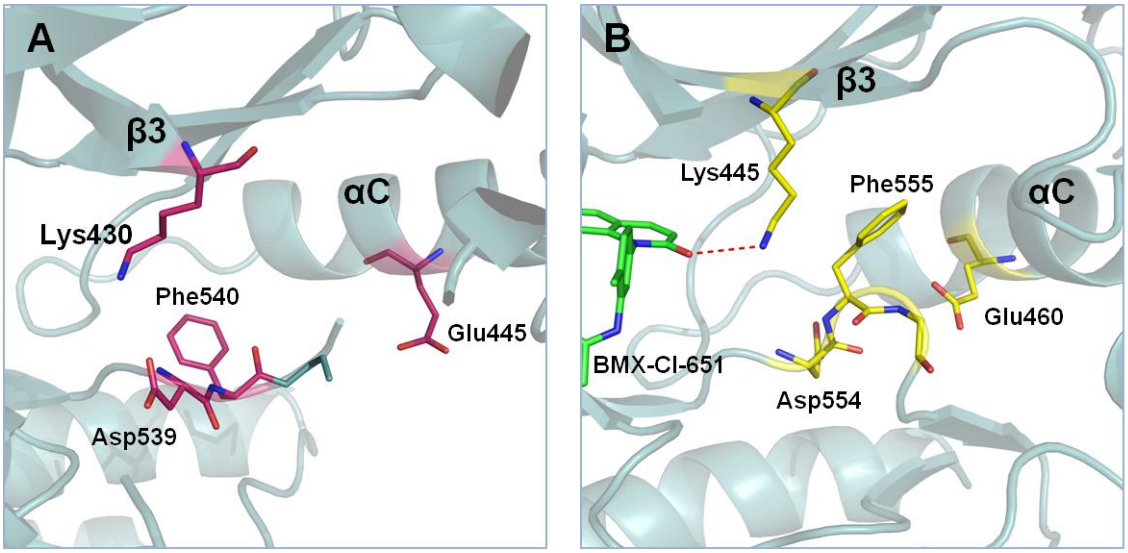

Figure 3.17. Positioning of the DFG Aspartic acid and Phenylalanine in the active and inactive kinase conformations. A) Close-up of the DFG Aspartic acid and Phenylalanine in apo BTK (PDB entry: 3P08). The DFG Asp540 points at the ATP binding pocket and the DFG Phe445 residue is positioned in the back cleft and facilitates the formation of the β3 Lys430-αC Glu455 ion pair. B) Close-up of the DFG Aspartic acid and Phenylalanine in the hBMX/BMX-CI-561 complex. The DFG Asp554 points down and away from the ATP binding pocket and the Phe555 is packed against the αC-helix and swings up, blocking the ion pair formation. The Lys445 also interacts with BMX-CI-651.

3.5.3. Interactions of BMX-CI-651 with hBMX catalytic domain

To counter the pathological role of BMX in prostate cancer, a series of BMX-IN-1 analogue molecules, inhibitors of BMX, were developed. BMX-IN-1 as an irreversible BMX inhibitor covalently modifies Cys496 located at the lip of the ATP binding site. The specificity of such positioning is critical to achieve selectivity against BMX, and is mostly modulated via a Michael acceptor motif.[28] However, although BMX kinase bears a cysteine 496 residue near its ATP site which is tremendously advantageous for the design of irreversible inhibitors, cysteine residues located at an identical position are also found in 11 other kinases such as TEC kinase family,

EGFR family, as well as BLK, JAK3, and MAP2K7 kinases. [101,102] Therefore, in order to achieve a high selectivity profiling against BMX, other complementary factors need to be introduced for targeting of noncysteine residues.[101,102]

Similar to BMX-IN-1, BMX-CI-651 is bound in the ATP cleft pocket and covalent inhibition of hBMX occurs via alkylation of the nucleophilic thiol group of Cys496 (Figure 3.18). This direct inhibition of hBMX is modulated via a Michael acceptor motif, which is characterized by an electrophilic moiety close to a double bond rendering the molecule highly active for nucleophilic attacks.[103] Other major polar non-binding interactions occur between the oxygen substituent in the pyridinone ring of BMX-CI-651 and the $–NH_2$ group of the residue Lys445, as well as, between the nitrogen in the BMX-CI-651 quinoline and the –NH of the residue Ile492 (Figure 3.18).

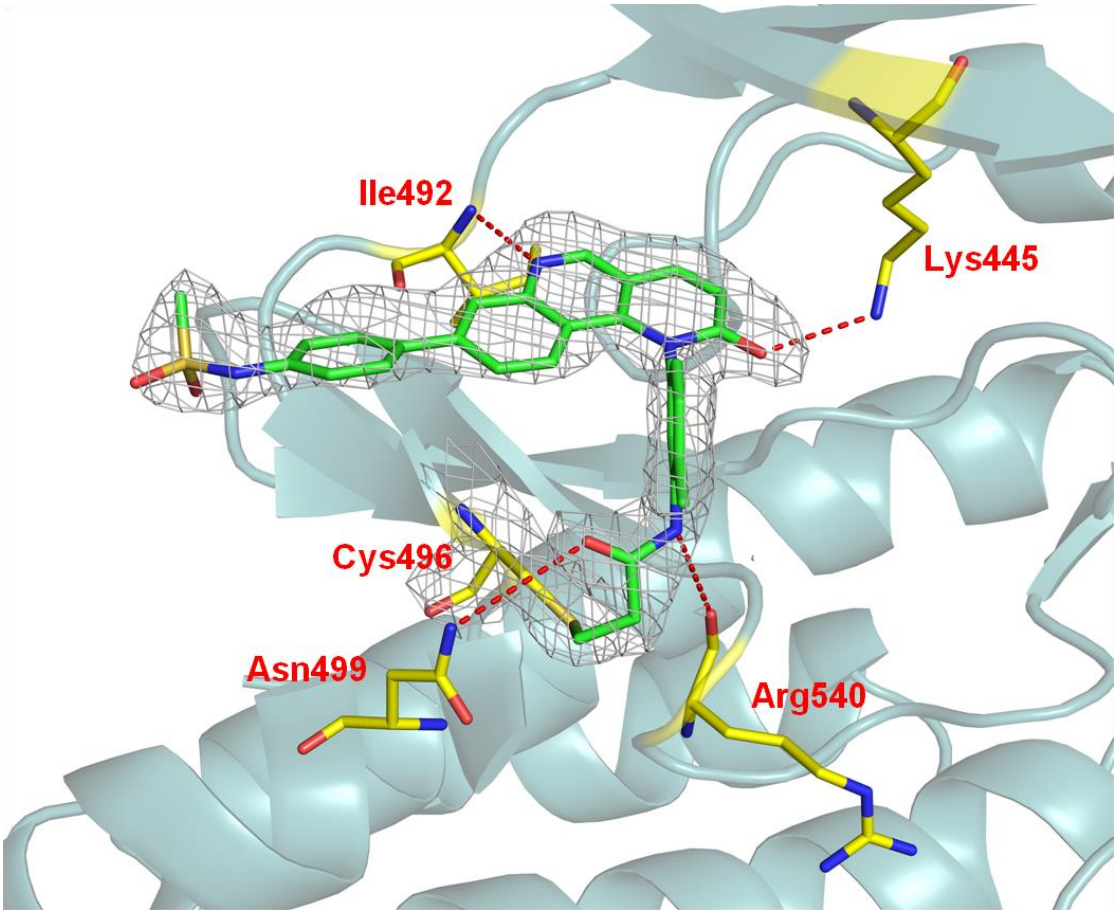

Figure 3.18. Binding mode of BMX-CI-651 to hBMX kinase catalytic domain. The covalent bond between the BMX-CI-651 and the nucleophilic Cys496 is showed as a yellow solid line. Other polar (non-binding) interactions of BMX-CI-651 with the residues Lys445, Ile492, Asn499, and Arg540 are represented as red dashed lines.

The polar interaction between BMX-CI-651 and Lys445 is one of the key points to regulate BMX activity. As previously mentioned, the conserved β3 Lys interacts with αC-helix Glu residue in order to form a salt bridge required for ATP hydrolysis. The binding of BMX-CI-651 to Lys445 alters this interaction and consequently inactivates BMX.

Other hydrophobic interactions occur between the aromatic rings of BMX-CI-651 and the side chains of Tyr491, Ala443, Val431, and Leu543 (Figure 3.19). BMX-CI-651 is further stabilized by a hydrogen bond between a water molecule and the carbonyl oxygen of the acrylamide group. A second water molecule stabilizes this first one via hydrogen bond and also forms a water bridge with the peptide nitrogen of Cys496 and the terminal amine group of Asn499.

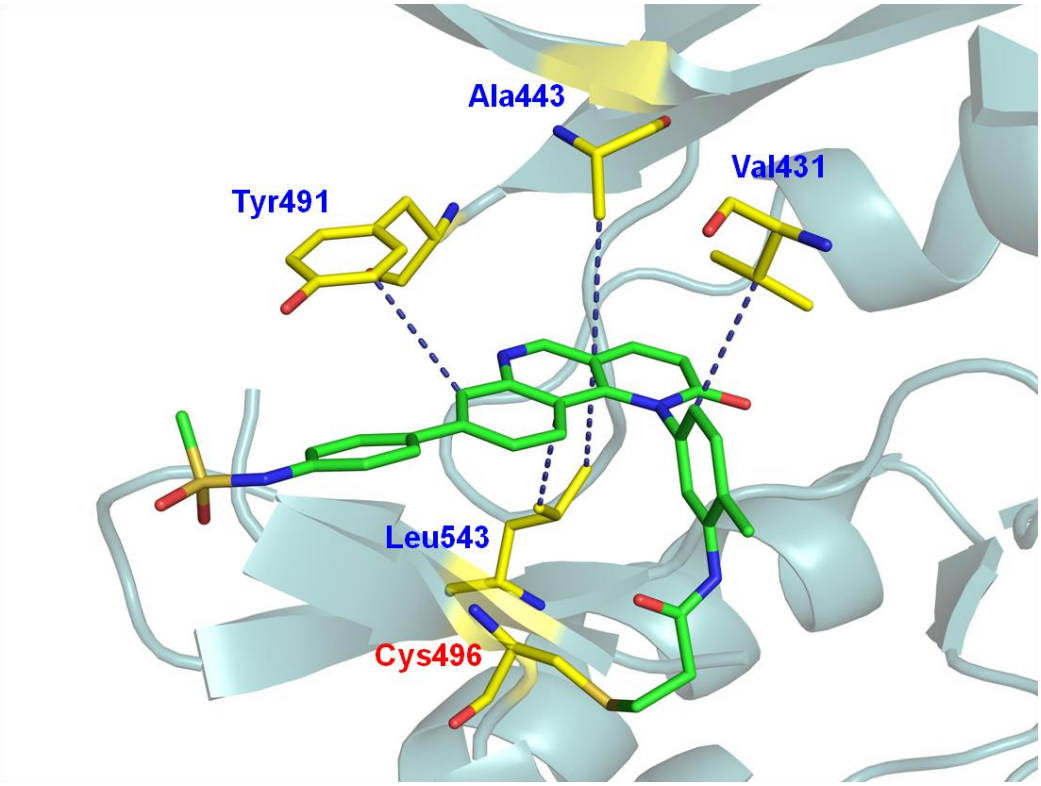

Figure 3.19. Hydrophobic interactions of the inhibitor BMX-CI-651 within the hBMX catalytic domain. The interactions with the residues Tyr491, Ala443, Val431, and Leu543 are represented as blue dashed lines.

The positioning of the BMX-CI-651 sulphonamide group is also of critical importance, since it is pointing outside of the ATP binding pocket and it does not seem to clash with any protein residues. Therefore, this “free tail” could be used for adding a linker or chemical handle for drug conjugation or even for imaging purposes.

The crystal structures of BMX in complex with BMX-CI-51 and BMX-CI-61 were not determined, due to lack of time. There is a possibility that these compounds present other advantageous binding properties which would allow them to achieve greater specificity towards BMX tyrosine kinase. This could be the key factor to differentiate BMX from other kinases that also bear a conserved cysteine residue adjacent to the ATP binding pocket, and as a result, selectively target BMX for prostate cancer therapy.

Chapter 4

4. Conclusions and future perspectives

Irreversible kinase inhibitors are highly desirable reagents used to permanently disable kinase activity, offering a number of pharmacological advantages over their reversible counterparts, including high potency, high selectivity, and prolonged pharmacodynamics for use as potential therapeutics.[102,104,105]

Despite their great utility, only a small number of well-characterized covalent inhibitors have received regulatory approval. This is mostly due to the high sequence conservation of the ATP binding determinants on the catalytic center and to the very large number of existing kinases, making the achievement of kinase selectivity a very challenging task. Therefore, new approaches are required for a rapid generation of kinase inhibitors that can be used as regulatory probes to specifically inhibit enzymatic activity.[102,104]

This master thesis plan was aimed at elucidating the binding mode of covalent inhibitors to the BMX protein, with the goal of improving the design and development of new inhibitors with increased potency and specificity. Structure determination by X-ray crystallography provided critical information to help understand the molecular basis of BMX inhibition.

The affinity and binding kinetics of 4 different covalent ligands to the hBMX kinase catalytic domain was analyzed, including BMX-IN-1 and three others that were developed based on the chemical structure of BMX-IN-1: BMX-CI-651, BMX-CI-51, and BMX-CI-61. In DSF experiments, BMX-CI-651, BMX-CI-51, and BMX-CI-61 showed a greater capability to stabilize the structure of hBMX in comparison with BMX-IN-1. The average melting temperature of hBMX with BMX-IN-1 was ca. 8ºC, but increased more than 9ºC in the presence of BMX-CI-651, BMX-CI-51, and BMX-CI-61. The K_D values calculated from SPR experiments, also showed that all newly designed compounds, BMX-CI-651, BMX-CI-51, and BMX-CI-61 have a greater affinity and inhibitory capabilities for hBMX kinase than BMX-IN-1.

Also, hBMX protein was expressed via an insect-baculovirus expression system with a satisfactory yield of 15.4 mg of protein per liter of culture. The purified sample was submitted to biochemical and biophysical characterization by SDS-PAGE, Western-Blot, Dynamic Light Scattering, and Circular Dichroism, allowing an assessment of the protein sample quality before the crystallization. Crystallization trials were set-up by co-crystallizing 6xHis-hBMX with BMX-CI-651 in order to generate crystals of sufficient quality to determine the three-dimensional structure of hBMX/BMX-CI-651 complex by X-ray crystallography.

The crystal structure of hBMX catalytic domain in complex with BMX-CI-651 was essential to get a molecular view of the ATP binding pocket and to understand how it binds to irreversible inhibitors. The results showed that not only the Michael acceptor moiety is important for

molecular binding, but also other substituents (e.g., in the pyridinone ring) contribute to non-covalent interactions. This combination of both covalent and noncovalent interactions is a key element to achieve a higher selectivity against BMX tyrosine kinase.

Hopefully, the determination of the 3D structure of the human BMX in complex with covalent inhibitors will play a key role for guiding further studies aiming to develop therapeutically active molecules. These drug design and discovery studies in combination with cell-based and *in vivo* testing, will provide a fundamental based characterization to explore a new chemical space, in order to effectively obtain lead compounds that can help overcome some of the current limitations encountered in cancer therapy.[106]

Acknowledgments

Before anything else, I would like to express my enormous gratitude to all the people who directly or indirectly contributed to this thesis project. Also, to the hosting entities and collaborators, ITQB-NOVA, iMM, and iBET, for the excellent working conditions.

To my supervisor Dr. Pedro M. Matias, for letting me join this project and giving me the opportunity to work at the Macromolecular Crystallography Unit. For the vote of confidence and respect that enable me to always push forward and to grow as a scientist.

To my co-supervisor, Dr. Marta C. Marques, for not only trusting and respecting my work, but also for providing me a rich source of inspiration as a successful woman in science. For the scientific and personal advices I will always carry with me throughout my professional journey.

To our collaborators at iBET, Dr. Paula Alves and Dr. António Roldão, for the amazing opportunity to join the Animal Cell Technology Unit (ACTU). Particularly, to Marcos Sousa, for the work and emotional support in the toughest times and for promoting my independence in the laboratory. For the example of determination and motivation in the workplace.

To Dr. Gonçalo Bernardes and Dr. João Seixas at GBernardes Lab (iMM), for providing me with all the funds and tools necessary for me to complete my work. Also, to all my colleagues, especially Maria Inês Albuquerque, for her truly and honest companionship, and support. I am most thankful for meeting all of you.

To Isabel Pacheco, Mónica Thomaz, Dr. Tiago Bandeiras, Dr. Pedro Sousa, Micael Freitas, Ana Rita Lemos, for helping me during the purification process and the biophysical characterization of BMX.

To Dr. José Brito, for the enthusiastic discussions that enable me to enrich my work and broadening my scientific perspectives.

To my colleagues at the Macromolecular Crystallography Unit for the good working environment and great times shared throughout this year. In particular, to Dr. Sara Silva and Diana Silva, for work support and friendship.

To my close friends Isabel Furtado, Maria Venâncio, Marta Silva, Inês Saraiva, Alexandra Teixeira, and Ricardo Barras, for making my life away from home much, much easier.

Aos meus irmãos, António e Tiago, e respectivas senhoras, Raquel Gravito e Carina Ventura, por todo apoio que sempre demonstraram e por tantos bocadinhos bons de felicidade. Adoro-vos de coração.

Por último, mas não menos importante, aos meus adorados pais, António e Maria, que sem o vosso apoio nunca teria chegado onde cheguei. Estou-vos eternamente agradecida.

Bibliography

1. Torre, L. A. *et al.* Global Cancer Statistics, 2012. *CA a cancer J. Clin.* **65,** 87–108 (2015).
2. Siegel, R., Miller, K. & Jemal, A. Cancer statistics , 2015 . *CA Cancer J Clin* **65,** 29 (2015).
3. Zhou, M. High-grade prostatic intraepithelial neoplasia, PIN-like carcinoma, ductal carcinoma, and intraductal carcinoma of the prostate. *Mod. Pathol.* **31,** S71-79 (2018).
4. Dong, J. Chromosomal deletions and tumor suppressor genes in prostate cancer. *Cancer Metastasis Rev. 2001;20(3-4)173-93.* **20,** 173–193 (2001).
5. Li, L. C., Carroll, P. R. & Dahiya, R. Epigenetic changes in prostate cancer: Implication for diagnosis and treatment. *J. Natl. Cancer Inst.* **97,** 103–115 (2005).
6. Gui, S. liang *et al.* Overexpression of CXCL3 can enhance the oncogenic potential of prostate cancer. *Int. Urol. Nephrol.* **48,** 701–709 (2016).
7. Abrahamsson, P. A. Neuroendocrine cells in tumor growth of the prostate. *Endocr. Relat. Cancer* **6,** 503–519 (1999).
8. Montgomery, R. B. *et al.* Maintenance of intratumoral androgens in metastatic prostate cancer: A mechanism for castration-resistant tumor growth. *Cancer Res.* **68,** 4447–4454 (2008).
9. Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: An evolving paradigm. *Nat. Rev. Cancer* **13,** 714–726 (2013).
10. Alfarouk, K. O. *et al.* Resistance to cancer chemotherapy: Failure in drug response from ADME to P-gp. *Cancer Cell Int.* **15,** 1–13 (2015).
11. Tamagnone, L. *et al.* BMX, a novel nonreceptor tyrosine kinase gene of the BTK/ITK/TEC/TXK family located in chromosome Xp22.2. *Oncogene* (1994).
12. Cenni, B., Gutmann, S. & Gottar-Guillier, M. BMX and its role in inflammation, cardiovascular disease, and cancer. *Int. Rev. Immunol.* **31,** 166–173 (2012).
13. Qiu, Y., Robinson, D., Pretlow, T. G. & Kung, H. J. Etk/Bmx, a tyrosine kinase with a pleckstrin-homology domain, is an effector of phosphatidylinositol 3'-kinase and is involved in interleukin 6-induced neuroendocrine differentiation of prostate cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* (1998). doi:10.1073/pnas.95.7.3644
14. Zhang, J., Yang, P. L. & Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9,** 28–39 (2009).
15. Muckelbauer, J. *et al.* X-ray crystal structure of bone marrow kinase in the X chromosome: A Tec family kinase. *Chem. Biol. Drug Des.* **78,** 739–748 (2011).
16. Zuccotto, F., Ardini, E., Casale, E. & Angiolini, M. Through the 'gatekeeper door': Exploiting the active kinase conformation. *Journal of Medicinal Chemistry* (2010). doi:10.1021/jm901443h
17. Weil, D., Power, M. a, Smith, S. I. & Li, C. L. Predominant expression of murine Bmx tyrosine kinase in the granulo-monocytic lineage. *Blood* (1997).
18. Smith, C. I. E. *et al.* The Tec family of cytoplasmic tyrosine kinases: Mammalian Btk,

Bmx, Itk, Tec, Txk and homologs in other species: Review articles. *BioEssays* (2001). doi:10.1002/bies.1062

19. Zhang, R. *et al.* Etk/Bmx Transactivates Vascular Endothelial Growth Factor 2 and Recruits Phosphatidylinositol 3-Kinase to Mediate the Tumor Necrosis Factor-induced Angiogenic Pathway. *J. Biol. Chem.* (2003). doi:10.1074/jbc.M310678200
20. Chen, R. *et al.* Regulation of the PH-domain-containing tyrosine kinase Etk by focal adhesion kinase through the FERM domain. *Nat. Cell Biol.* (2001). doi:10.1038/35074500
21. Tsai, Y. T. *et al.* Etk, a Btk family tyrosine kinase, mediates cellular transformation by linking Src to STAT3 activation. *Mol. Cell. Biol.* (2000). doi:10.1128/MCB.20.6.2043-2054.2000
22. Guo, Z. *et al.* Regulation of androgen receptor activity by tyrosine phosphorylation. *Cancer Cell* (2006). doi:10.1016/j.ccr.2006.08.021
23. Mahajan, N. P. *et al.* Activated Cdc42-associated kinase Ack1 promotes prostate cancer progression via androgen receptor tyrosine phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* (2007). doi:10.1073/pnas.0700420104
24. Dai, B. *et al.* Compensatory upregulation of tyrosine kinase Etk/BMX in response to androgen deprivation promotes castration-resistant growth of prostate cancer cells. *Cancer Res.* **70,** 5587–5596 (2010).
25. Fox, J. L. & Storey, A. BMX negatively regulates BAK function, thereby increasing apoptotic resistance to chemotherapeutic drugs. *Cancer Res.* (2015). doi:10.1158/0008-5472.CAN-14-1340
26. Jiang, T. *et al.* Bi-directional regulation between tyrosine kinase Etk/BMX and tumor suppressor p53 in response to DNA damage. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M409108200
27. Jarboe, J. S., Dutta, S., Velu, S. E. & Willey, C. D. Mini-Review: Bmx Kinase Inhibitors for Cancer Therapy. *Recent Pat. Anticancer. Drug Discov.* **8,** 228–238 (2013).
28. Liu, F. *et al.* Discovery of a selective irreversible BMX inhibitor for prostate cancer. *ACS Chem. Biol.* **8,** 1423–1428 (2013).
29. Chen, X.-L., Qiu, L., Wang, F. & Liu, S. Current understanding of tyrosine kinase BMX in inflammation and its inhibitors. *Burn. Trauma* (2014). doi:10.4103/2321-3868.135483
30. Contreras-Gómez, A., Sánchez-Mirón, A., García-Camacho, F., Molina-Grima, E. & Chisti, Y. Protein production using the baculovirus-insect cell expression system. *Biotechnol. Prog.* (2014). doi:10.1002/btpr.1842
31. Min, L. *et al.* Disrupting the Intermolecular Self-Association of Itk Enhances T Cell Signaling. *J. Immunol.* (2010). doi:10.4049/jimmunol.0901908
32. Miller, L. K. Introduction to the Baculoviruses. in *The Baculoviruses* (1997). doi:10.1007/978-1-4899-1834-5_1
33. Vicente, T., Roldão, A., Peixoto, C., Carrondo, M. J. T. & Alves, P. M. Large-scale production and purification of VLP-based vaccines. *Journal of Invertebrate Pathology*

(2011). doi:10.1016/j.jip.2011.05.004

34. Rohrmann, G. The baculovirus replication cycle : Effects on cells and insects. *Baculovirus Mol. Biol.* (2013).
35. Peng, K., van Oers, M. M., Hu, Z., van Lent, J. W. M. & Vlak, J. M. Baculovirus Per Os Infectivity Factors Form a Complex on the Surface of Occlusion-Derived Virus. *J. Virol.* (2010). doi:10.1128/JVI.00812-10
36. Luckow, V. a, Lee, S. C., Barry, G. F. & Olins, P. O. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. *J. Virol.* (1993).
37. Wu, T. Y., Chen, Y. J., Teng, C. Y., Chen, W. S. & Villaflores, O. A bi-cistronic baculovirus expression vector for improved recombinant protein production. *Bioeng Bugs* (2012). doi:10.4161/bbug.19388
38. Miele, S. A. B., Garavaglia, M. J., Belaich, M. N. & Ghiringhelli, P. D. Baculovirus: Molecular Insights on Their Diversity and Conservation. *Int. J. Evol. Biol.* (2011). doi:10.4061/2011/379424
39. Wang, X., Ooi, B. G. & Miller, L. K. Baculovirus vectors for multiple gene expression and for occluded virus production. *Gene* (1991). doi:10.1016/0378-1119(91)90358-I
40. Vaughn, J. L., Goodwin, R. H., Tompkins, G. J. & McCawley, P. The establishment of two cell lines from the insectspodoptera frugiperda (lepidoptera; noctuidae). *In Vitro* (1977). doi:10.1007/BF02615077
41. O'Reilly, D. R., Miller, L. K. & Luckow, V. A. Baculovirus Expression Vectors: A Laboratory Manual. *Oxford Univ. Press. New York u.a.* (1994).
42. Kato, T., Kajikawa, M., Maenaka, K. & Park, E. Y. Silkworm expression system as a platform technology in life science. *Applied Microbiology and Biotechnology* (2010). doi:10.1007/s00253-009-2267-2
43. Invitrogen. Bac-to-Bac® Baculovirus Expression System. *User Guid.* 1–78 (2015). doi:10.1007/SpringerReference_28001
44. Neermann, J. & Wagner, R. Comparative analysis of glucose and glutamine metabolism in transformed mammalian cell lines, insect and primary liver cells. *J. Cell. Physiol.* (1996). doi:10.1002/(SICI)1097-4652(199601)166:1<152::AID-JCP18>3.0.CO;2-H
45. Roldão, A., Carrondo, M. J. T., Alves, P. M. & Oliveira, R. Stochastic simulation of protein expression in the baculovirus/insect cells system. *Comput. Chem. Eng.* (2008). doi:10.1016/j.compchemeng.2007.04.017
46. Roldão, A., Oliveira, R., Carrondo, M. J. T. & Alves, P. M. Error assessment in recombinant baculovirus titration: Evaluation of different methods. *J. Virol. Methods* **159,** 69–80 (2009).
47. Figliozzi, R. W., Chen, F., Chi, A. & Hsia, S. C. V. Using the inverse Poisson distribution to calculate multiplicity of infection and viral replication by a high-throughput fluorescent imaging system. *Virologica Sinica* (2016). doi:10.1007/s12250-015-3662-8
48. Mena, J. A., Ramírez, O. T. & Palomares, L. A. Titration of non-occluded baculovirus

using a cell viability assay. *Biotechniques* (2003).

49. Stetefeld, J., McKenna, S. A. & Patel, T. R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical Reviews* (2016). doi:10.1007/s12551-016-0218-6
50. Borgstahl, G. E. O. How to use dynamic light scattering to improve the likelihood of growing macromolecular crystals. *Methods Mol. Biol.* (2007). doi:10.1385/1-59745-209-2:109
51. Proteau, A., Shi, R. & Cygler, M. Application of dynamic light scattering in protein crystallization. *Curr. Protoc. Protein Sci.* (2010). doi:10.1002/0471140864.ps1710s61
52. Miles, A. J. & Wallace, B. A. Circular Dichroism Spectroscopy for Protein Characterization: Biopharmaceutical Applications. in *Biophysical Characterization of Proteins in Developing Biopharmaceuticals* (2014). doi:10.1016/B978-0-444-59573-7.00006-3
53. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751,** 119–139 (2005).
54. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* (2006). doi:10.1038/nprot.2006.202.Using
55. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* (2007). doi:10.1038/nprot.2007.321
56. Naik, S., Zhang, N., Gao, P. & Fisher, M. T. On the design of broad based screening assays to identify potential pharmacological chaperones of protein misfolding diseases. in *Current topics in medicinal chemistry* (2012). doi:10.2174/1568026611212220006
57. Fedorov, O., Niesen, F. H. & Knapp, S. Kinase inhibitor selectivity profiling using differential scanning fluorimetry. *Methods Mol. Biol.* (2012). doi:10.1007/978-1-61779-337-0_7
58. Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* (2006). doi:10.1016/j.ab.2006.07.027
59. Jena Bioscience GmbH. JBScreen Thermofluor. Available at: https://www.jenabioscience.com/crystallography-cryo-em/screening/solubility-screens/jbscreen-thermofluor. (Accessed: 31st July 2018)
60. Nguyen, H. H., Park, J., Kang, S. & Kim, M. Surface plasmon resonance: A versatile technique for biosensor applications. *Sensors (Switzerland)* (2015). doi:10.3390/s150510481
61. Ciulli, A. Biophysical screening for the discovery of small-molecule ligands. *Methods Mol. Biol.* (2013). doi:10.1007/978-1-62703-398-5_13
62. GE Healthcare Life Sciences. Biacore Assay Handbook. *GE Healthc. Bio-Sciences AB* (2012).
63. Frostell, Å., Vinterbäck, L. & Sjöbom, H. Protein-Ligand Interactions Using SPR

Systems. *Methods Mol. Biol.* (2013). doi:10.1007/978-1-62703-398-5_6

64. Rhodes, G. *Crystallography Made Crystal Clear (Third Edition)*. *Aerospace medicine* (2002). doi:10.1016/B978-012587073-3/50001-5
65. Glazer, A. M. & Burns, G. *Space Groups for Solid State Scientists*. *Space Groups for Solid State Scientists* (2013). doi:10.1016/C2011-0-05712-5
66. Glusker, J. P. ., Lewis, M. & Rossi, M. Crystal Structure Analysis for Chemists and Biologists. *VCH Publishers* (1994). doi:10.1021/ed072pA73.9
67. Krauss, I. R., Merlino, A., Vergara, A. & Sica, F. An overview of biological macromolecule crystallization. *International Journal of Molecular Sciences* (2013). doi:10.3390/ijms140611643
68. Asherie, N. Protein crystallization and phase diagrams. *Methods* (2004). doi:10.1016/j.ymeth.2004.03.028
69. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallographica Section F:Structural Biology Communications* (2014). doi:10.1107/S2053230X13033141
70. Snell, E. H. & Helliwell, J. R. Macromolecular crystallization in microgravity. *Reports Prog. Phys.* (2005). doi:10.1088/0034-4885/68/4/R02
71. Danley, D. E. Crystallization to obtain protein-ligand complexes for structure-aided drug design. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2006). doi:10.1107/S0907444906012601
72. Hassell, A. M. *et al.* Crystallization of protein-ligand complexes. in *Acta Crystallographica Section D: Biological Crystallography* (2006). doi:10.1107/S0907444906047020
73. Hofbauer, S. *et al.* Stabilization of porcine pancreatic elastase crystals by glutaraldehyde cross-linking. *Acta Crystallogr. Sect. Struct. Biol. Commun.* (2015). doi:10.1107/S2053230X15017045
74. Smyth, M. S. & Martin, J. H. J. x Ray crystallography. *Journal of Clinical Pathology - Molecular Pathology* (2000). doi:10.1136/mp.53.1.8
75. Taylor, G. L. Introduction to phasing. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/S0907444910006694
76. Ciccarone, V. C., Polayes, D. a & Luckow, V. a. Generation of Recombinant Baculovirus DNA in E.coli Using a Baculovirus Shuttle Vector. *Methods Mol. Med.* (1998). doi:10.1385/0-89603-485-2:213
77. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2011). doi:10.1107/S0907444911007773
78. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* (2007). doi:10.1107/S0021889807021206
79. Potterton, L. *et al.* Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2004). doi:10.1107/S0907444904023716
80. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* (2003).

doi:10.1107/S0907444903008126

81. Cowtan, K. The Buccaneer software for automated model building. *Acta Crystallogr D Biol Crystallogr* (2006). doi:10.1107/S0907444906022116
82. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2004). doi:10.1107/S0907444904019158
83. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* (1997). doi:10.1107/S0907444996012255
84. Lebedev, A. A. *et al.* JLigand: A graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2012). doi:10.1107/S090744491200251X
85. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/S0907444909052925
86. Painter, J. & Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* (2006). doi:10.1107/S0021889805038987
87. DeLano, W. L. The PyMOL Molecular Graphics System. *Schrödinger LLC wwwpymolorg* (2002). doi:citeulike-article-id:240061
88. Davis, I. W. *et al.* MolProbity : all-atom contacts and structure validation for proteins and nucleic acids. **35,** 375–383 (2007).
89. Üstün-Aytekin, Ö., Gürhan, I. D., Ohura, K., Imai, T. & Öngen, G. Monitoring of the effects of transfection with baculovirus on Sf9 cell line and expression of human dipeptidyl peptidase IV. *Cytotechnology* (2014). doi:10.1007/s10616-013-9549-3
90. Radner, S. *et al.* Transient transfection coupled to baculovirus infection for rapid protein expression screening in insect cells. *J. Struct. Biol.* (2012). doi:10.1016/j.jsb.2012.04.013
91. Sandhu, K. S., Naciri, M. & Al-Rubeai, M. Prediction of recombinant protein production in an insect cell-baculovirus system using a flow cytometric technique. *J. Immunol. Methods* (2007). doi:10.1016/j.jim.2007.06.005
92. Radford, K. M. *et al.* The indirect effects of multiplicity of infection on baculovirus expressed proteins in insect cells: Secreted and non-secreted products. *Cytotechnology* (1997). doi:10.1023/A:1007962903435
93. Nishikawa, N., Yamaji, H. & Fukuda, H. Recombinant protein production by the baculovirus-insect cell system in basal media without serum supplementation. in *Cytotechnology* (2003). doi:10.1023/B:CYTO.0000039894.27256.0f
94. Hahnefeld, C., Drewianka, S. & Herberg, F. W. Determination of kinetic data using surface plasmon resonance biosensors. *Methods Mol. Med.* (2004). doi:10.1385/1-59259-679-7:299
95. Holdgate, G. A., Anderson, M., Edfeldt, F. & Geschwindner, S. Affinity-based, biophysical methods to detect and analyze ligand binding to recombinant proteins: Matching high information content with high throughput. *J. Struct. Biol.* (2010). doi:10.1016/j.jsb.2010.06.024

96. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A* (1976). doi:10.1107/S0567739476001873

97. Mao, C., Zhou, M. & Uckun, F. M. Crystal structure of Bruton's tyrosine kinase domain suggests a novel pathway for activation and provides insights into the molecular basis of X-linked agammaglobulinemia. *J. Biol. Chem.* (2001). doi:10.1074/jbc.M104828200

98. Liao, J. J.-L. Molecular Recognition of Protein Kinase Binding Pockets for Design of Potent and Selective Kinase Inhibitors. *J. Med. Chem.* (2007). doi:10.1021/jm0608107

99. Brown, N. R., Noble, M. E. M., Endicott, J. A. & Johnson, L. N. The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nat. Cell Biol.* (1999). doi:10.1038/15674

100. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* (2002). doi:10.1016/S0092-8674(02)00741-9

101. Li, X. *et al.* Discovery of a series of 2,5-diaminopyrimidine covalent irreversible inhibitors of Bruton's tyrosine kinase with in vivo antitumor activity. *J. Med. Chem.* (2014). doi:10.1021/jm4017762

102. Liang, X. *et al.* Discovery of 2-((3-Acrylamido-4-methylphenyl)amino)-N-(2-methyl-5-(3,4,5-trimethoxybenzamido)phenyl)-4-(methylamino)pyrimidine-5-carboxamide (CHMFL-BMX-078) as a Highly Potent and Selective Type II Irreversible Bone Marrow Kinase in the X Chromosome (BMX). *J. Med. Chem.* **60,** 1793–1816 (2017).

103. Maucher, I. V. *et al.* Michael acceptor containing drugs are a novel class of 5-lipoxygenase inhibitor targeting the surface cysteines C416 and C418. *Biochem. Pharmacol.* (2017). doi:10.1016/j.bcp.2016.11.004

104. Liu, Q. *et al.* Developing Irreversible Inhibitors of the Protein Kinase Cysteinome. *Chem. Biol.* (2013). doi:10.1016/j.chembiol.2012.12.006

105. Ferguson, F. M. & Gray, N. S. Kinase inhibitors: The road ahead. *Nature Reviews Drug Discovery* (2018). doi:10.1038/nrd.2018.21

106. Jaroch, K., Jaroch, A. & Bojko, B. Cell cultures in drug discovery and development: The need of reliable in vitro-in vivo extrapolation for pharmacodynamics and pharmacokinetics assessment. *Journal of Pharmaceutical and Biomedical Analysis* (2018). doi:10.1016/j.jpba.2017.07.023