Development and characterisation of lanthanide complexes for structural biology

Zaynab Alsalman

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : PHYS - Physique
Spécialité : Physique pour les Sciences du Vivant
Unité de recherche : Institut de Biologie Structurale

Development and characterisation of lanthanide complexes for structural biology.

Développement et caractérisation de complexes de lanthanide pour la biologie structurale.

Présentée par :

ZAYNAB ALSALMAN

Direction de thèse :

Eric GIRARD DIRECTEUR DE RECHERCHE, CEA Grenoble (Institut de Biologie Structurale, Université Grenoble Alpes)	Directeur de thèse

Thèse soutenue publiquement le **14/04/2023**,

devant le jury composé de :

Véronique RECEVEUR-BRECHOT DIRECTRICE DE RECHERCHE, CNRS Provence, BIP (Université d'Aix-Marseille, UMR 7281)	Rapporteure
Ahmed HAOUZ INGENIEUR DE RECHERCHE, Institut Pasteur de Paris	Rapporteur
Nicolas GIRAUD PROFESSEUR, Université Paris Descartes	Examinateur
Claude SAUTER DIRECTEUR DE RECHERCHE, CNRS Alsace, IBMC (Université de Strasbourg)	Président
Monika SPANO MAITRE DE CONFERENCES, Université Grenoble Alpes (Institut de Biologie Structurale)	Examinatrice

Acknowledgments

First of all, I would like to sincerely thank the members of my thesis jury for agreeing to assess and correct my manuscript and for their valuable time, expertise, and insightful feedback during the evaluation process.

I would like to express my deepest gratitude to my supervisor, Eric Girard, for your support, patience, guidance and mentorship throughout this journey. I am really thankful to have had you as my supervisor.

I would like to express my gratitude to the collaborators at ENS Lyon, who have played a significant role in the completion of my PhD thesis, Amandine Roux-Gossart, Natacha Gillet, François Riobé, Elise Dumont and Olivier Maury.

I would like to extend my appreciation to ELMA Team Bruno , Frank, Domi and Mylène for your kindness and encouragement. Also, I would like to thank Jacques Covès for his support, his contribution in my thesis and his scientific comments and advices that I had always appreciated.

I am also grateful to my colleagues, Adeline, Jennifer, Emilie, Quentin and Laura for their kindness, endless support, advices and for all the good moments and discussion that we had.

To my colleague and friend Lorenzo, I would like to express my sincere appreciation for all the countless conversations and for your unwavering support, encouragement and belief in my abilities.

To my friends Alessia, Ghina, Jamal, Nancy, Farah, Kadi, Myriam, Antonia, Mayssam, Leila, Rida, Rime and Zaynab and to my dear cousins in France, Hassan, Zeinab, Fatme and Nour thank you for your unwavering support, encouragement, and understanding throughout this journey.

I am eternally grateful to my family for their unwavering love, encouragement, and belief in my abilities. To my source of happiness, my nephew and nieces, Ali M, Ali S, Hassouna, Lyne and Serena, my siblings, Balsam, Mohamad and Shahnaz, and especially my Mom, thank you for being my pillars of strength. Your continuous support have been my driving force, and I am blessed to have you in my life.

To my dear Father, even though you are not physically present in this life, your spirit remains always alive in my heart. Your memory remind me to never give up and always strive for excellence, hoping to make you always proud.

Table of contents

List of Figures 6
List of Tables 10
List of Abbreviations 11

1. INTRODUCTION 13

1.1 Getting a quality protein sample 17
1.2 Crystallization theory 18
1.2.1 Phase diagram 18
1.2.2 Nucleation thermodynamic 20
1.3 Crystallization methods 24
1.3.1 Batch 25
1.3.2 Vapour Diffusion 26
1.3.3 Dialysis 26
1.3.4 Free-interface and counter-diffusion methods 27
1.4 Optimization of crystallization 28
1.4.1 Optimization of crystallization conditions (crystallization cocktail) 29
1.4.2 Seeding 31
1.4.3 Nucleating agents 32
1.4.3.1 Porous materials 33
1.4.3.2 Molecularly imprinted polymers (MIPs) 33
1.4.3.3 Silver Bullets 34
1.4.3.4 Molecular Glues 35
a) Polyoxometalates (POMs) 36
b) Calixarenes 37
1.5 Way to the third generation of Lanthanide Complexes 39
1.5.1 Lanthanide complexes : history → TbXo4 39
1.5.2 TbXo4 properties 40
1.5.2.1 Phasing properties 40
1.5.2.2 Nucleating Properties 41
1.5.2.3 TbXo4 binding mode 44
1.6 Aim of the thesis 46

2. MATERIALS AND METHODS 49

2.1 Preparation of commercial protein's solutions 49
2.2 Recombinant Proteins 49
2.2.1 Expression 49
2.2.2 Purification protocol of recombinant proteins (MthAdkA) 52
2.3 Purification protocol of Egg-White Lysozyme 55
2.4 Crystallization screening 57
2.5 Protein structure determination by X-ray crystallography 57
2.5.1 Diffraction data recording 57
2.5.2 Diffraction data processing 58

3. EVALUATION OF TBXO4 VARIANTS 62

3.1 TbXo4 reproducibility issues 62
3.2 TbXo4 variant description 74
3.2.1 Evaluation of variant nucleating properties 75
3.2.2 Presence of self-crystallization 76
3.2.3 Results with Hen Egg-White Lysozyme 76
3.2.4 Results with ProteinaseK 79
3.2.5 Results with Thaumatin 80
3.2.6 Conclusion on Xo4 variant nucleating properties. 82
3.3 Determination of TbXo4 and Xo4 variants binding mode in nucleating condition 82

3.3.1 Structure of HEWL with 10 mM TbXo4 ..... *83*
3.3.2 HEWL structures with selected Family2 variants *88*
3.3.3 QM/MM molecular simulations *95*
3.4 CONCLUSION 97

4. CHARACTERISATION AND UNDERSTANDING OF TBXO4 INTERACTIONS 99

4.1 MODEL 1: MTHADKA 100
4.1.1 MthAdkA native structure analysis and design of the mutants *100*
4.1.2 Crystallization output comparisons between WT/mutant MthAdkA. *105*
4.1.3 WT and mutant MthAdkA structure analysis *113*
4.1.4 MthAdkA conclusion *118*
4.2 MODEL 2: EGG-WHITE LYSOZYME (FROM THE EGGS TO THE STRUCTURES: EGG HUNT) 120
4.2.1 Effect of the natural mutations on the crystallization of EWL in the presence of TbXo4 ..... *122*
4.2.2 Evaluation of TbXo4 interaction with EWL *128*
4.2.3 Molecular Dynamic simulation (MM/QM) *136*
4.2.4 Conclusion EWL *138*

5. CONCLUSION AND PERSPECTIVES 141

6. RESUME DE LA THESE EN FRANÇAIS 145

6.1 INTRODUCTION 145
6.2 ÉVALUATION DES VARIANTS DE TBXO4 149
6.3 CARACTERISATION ET COMPREHENSION DU MECANISME D'INTERACTION DE TBXO4 151
6.3.1 Adénylate kinase *152*
6.3.2 Chasse aux œufs : Lysozyme de blanc d'œuf (EWL) *156*

REFERENCES 159

List of Figures

Figure 1-1 Number of released PDB structures per Year for each method (RSCB PDB Nov-2022) ... 14

Figure 1-2 X-Ray crystallography workflow for the determination of protein structure. ... 15

Figure 1-3 Statistics related to the structure determination process. From (Khurshid et al., 2014) ... 16

Figure 1-4 Protein crystallization phase diagram based on two variables, protein and precipitant concentrations. ... 19

Figure 1-5 The variation of the Gibbs Free energy (ΔG) according to particle radius (r) during nucleation phenomenon. ... 23

Figure 1-6 Protein crystallization phase diagram based on two variable parameters, protein and precipitant concentrations (N. E. Chayen, 1998) in four major crystallization methods: A. Batch, B. Vapour Diffusion, C. Dialysis and D. Free-interface diffusion. ... 24

Figure 1-7 Principles of major methods used to crystallize biological macromolecules (a) a version of batch crystallization. (b) Vapour-diffusion crystallization with sitting and hanging drops. (c) Dialysis method with Cambridge button. (d) Crystallization by free-interface diffusion in a capillary where two solutions of equivalent volume are brought into contact. (e) Counter-diffusion setup in X-ray capillary tubes pinched in a gel layer. (adapted from (Sauter et al., 2012)). ... 25

Figure 1-8 Evolution of the protein solubility (log S) as a function of the square root of ionic strength (I) of the crystallization solution (McPherson & Gavira, 2014). ... 30

Figure 1-9 Graphic showing the most frequent POM–protein interactions (Bijelic & Rompel, 2015) ... 36

Figure 1-10 Structure of calix[4]arene. R = $SO3-$, $HPO3-$, $CH2-HPO3-$ ... 37

Figure 1-11 Three generations of lanthanide complexes designed for protein structure determination (Engilberge et al., 2017) ... 39

Figure 1-12 Automated crystallization screening (576 conditions) performed on 8 proteins (drops observed after 34 days). grey: number of hits for native protein, dots: number of hits for protein supplemented with 10 mM of TbXo4 (Engilberge et al., 2017). ... 42

Figure 1-13 Crystallization diagrams determined for T5-phage pb6 protein in the absence (left) and in the presence (right) of TbXo4 (10 mM). Black dots represent conditions leading to crystals (hits). White dots represent clear drops. Crosses represent drops with protein precipitates. (b) Comparison of the crystals' quality obtained under strictly similar crystallization conditions for the two proteins of unknown structures pb6 and MDH, in the absence and presence of 10 mM of TbXo4 as indicated (Engilberge et al., 2017). ... 43

Figure 1-14 Samples obtained from M. thermolithotrophicus by native purification. (a) SDS-PAGE gels of the purified proteins fractions. (b) 96 Crystallization screening conditions performed on the proteins fractions. The number of unique crystallization hits is depicted in grey for the native protein without TbXo4 and with dots for the protein supplemented with 10 mM TbXo4. Noted that Fraction A and D were only evaluated with TbXo4. (c-f) Crystallization results on fractions A to D in presence of 10 mM TbXo4 (Engilberge et al., 2019). ... 44

Figure 1-15 Illustration of TbXo4 dimer interactions in HEWL structure. HEWL was co-crystallized with 100mM TbXo4 a) crystal packing of HEWL in presence of TbXo4 dimer. b) TbXo4 dimer and its interactions in the structure of HEWL. Asymmetric unit content are shown in green and symmetry-related molecules represented in light brown. Values in blue refer to interaction energies evaluated by DFT (in kcal/mol) PDB: 6F2I. (adapted from (Engilberge et al., 2018)). ... 45

Figure 2-1 Example of elution chromatogram of MthAdkA after an ion exchange column and its SDS-PAGE gel on the right. The blue and red curves represent the U.V. absorbance at 280 nm and 254 nm, respectively. SN_b: Supernatant before heat, SN_a: Supernatant after heat, FT: Flowthrough and PF: Protein Fractions. ... 53

Figure 2-2 Example of elution chromatogram of MthAdkA after gel filtration and its SDS-PAGE gel on the right. The blue and red curves represent the U.V. absorbance at 280 nm and 254 nm, respectively. Collected Protein fractions i.e. protein fractions used for crystallization. ... 54

Figure 2-3 Example of elution chromatogram of EWL (here QEWL) after gel filtration. The blue curve represents the U.V. absorbance at 280nm. Collected Protein fractions used for crystallization are indicated. 56

Figure 3-1 TbXo4 reproducibility test a) Crystalline precipitates of pb9 protein from phage T5 after mixing with 10 mM TbXo4 in the Eppendorf tube (left) and the microscopic view of the mixture (right) b) precipitate of catalase from bovine liver after mixing with TbXo4 to a final concentration of 1 0mM (left), clear drop after adding $HCO3-$(right). 63

Figure 3-2 Number of crystallization hits of three commercial proteins, HEWL (Hen Egg-White Lysozyme), ProteinaseK and Thaumatin in presence of TbXo4 batches, using 1+1 crystallization protocol test. In (blue) the average (avg) number of crystallization hits using the new TbXo4 batch and its standard deviation in red error bars and in (grey) the number of crystallization hits from TbXo4-SE batch (Engilberge et al., 2017). 64

Figure 3-3 Number of crystallization hits of HEWL (a) in absence blue and (b) in presence (orange) of TbXo4, using 1+1 crystallization protocol test, the average(avg) and its standard deviation(error bars in red) and cumulative number (cml) of all the assays. c) Origin of protein lot and TbXo4 batch for each crystallization assay. 65

Figure 3-4 Number of crystallization hits of ProteinaseK in absence (blue) and in presence (orange) of TbXo4, using 1+1 crystallization protocol test, the average(avg) and its standard deviation(error bars in red) and cumulative number (cml) of all the assays. c) Origin of protein lot and TbXo4 batch for each crystallization assay. 67

Figure 3-5 Number of crystallization hits of Thaumatin in absence (blue) and in presence (orange) of TbXo4, using 1+1 crystallization protocol test, the average(avg) and its standard deviation(error bars in red) and cumulative number (cml) of all the assays. c) Origin of protein lot and TbXo4 batch for each crystallization assay. 69

Figure 3-6 Representation and Subset Of Interest analysis (SOI Area tables) of the crystallization assays of six proteins against (a) commercial crystallization kit and (b) pH. Each square represents a crystallization hit. The three areas of SOI correspond to the unique hits observed for the proteins in the absence/presence of TbXo4 (area -1/+1, respectively) and observed in both cases (area 0). The pH value shown corresponds to the value presented in the crystallization conditions and available in the description of the commercial kits. None: indicates that no value is provided in the description (Jiang et al., 2020)..... 70

Figure 3-7 Illustration of the cumulative crystallization results from four crystallization identical assays for the same protein. Assays 1&2 and 3&4 resulted in 2 and 3 hits, respectively. The corresponding cumulative value is 4. 71

Figure 3-8 Number of crystallization hits of a) three commercial proteins, HEWL (Hen Egg-White Lysozyme), ProteinaseK and Thaumatin in presence of TbXo4 batches, using 1+1 crystallization protocol test. In blue the average (avg) number of crystallization hits using the new TbXo4 batch and its standard deviation in red error bars, in grey the number from TbXo4-SE batch (Engilberge et al., 2017) and in orange the number in presence of 1eq. $HCO3-$. b) HEWL, pb9 & LDH, using 1+1+1 crystallization protocol. In blue the number of crystallization hits in absence of $HCO3-$ and in orange the number in presence of 1eq. $HCO3-$. 73

Figure 3-9 Crystallophore variants produced by the team of Olivier Maury at ENS-Lyon. The three families are described within the text. The values between brackets indicate the electrostatic charge of the complex. 75

Figure 3-10 Results of the automated crystallization screening of crystallophore variants with HEWL. Drops observation after 30 days: number of unique hits for native protein in blue, common hits (native and Xo4) in orange, Xo4 variants unique hits in grey. 78

Figure 3-11 Comparison of variant versus TbXo4 crystallization outputs in the case of HEWL. Drop observation after 30 days: TbXo4 unique hits in blue, common hits (TbXo4 and Xo4 variants) in orange, Xo4 variants unique hits in grey. 78

Figure 3-12 Results of the automated crystallization screening of crystallophore variants with ProteinaseK. Drop observation after 30 days: number of unique hits for native protein in blue, common hits (native and Xo4) in orange, Xo4 variants unique hits in grey. 79

Figure 3-13 Comparison of variant versus TbXo4 crystallization output in the case of ProteinaseK. Drop observation after 30 days: TbXo4 unique hits in blue, common hits (TbXo4 and Xo4 variants) in orange,

Xo4 variants unique hits in grey. 80

Figure 3-14 Results of the automated crystallization screening of crystallophore variants with Thaumatin after 30 days: number of unique hits for native protein in blue, common hits (native and Xo4) in orange, Xo4 variants unique hits in grey. 81

Figure 3-15 Comparison of variant versus TbXo4 crystallization outputs in the case of Thaumatin. Drop observation after 30 days: TbXo4 unique hits in blue, common hits (TbXo4 and Xo4 variants) in orange, Xo4 variants unique hits in grey. 81

Figure 3-16 Structure of HEWL co-crystallised with 10 mM TbXo4. a) anomalous Fourier map (contour: 5 sigma). 2Fo-Fc electron density map (contour: 1 sigma) of b) TbXo4 binding residues and c&e) the TbXo4 molecule (two views). d&f) Interaction sites of TbXo4 (Two same views as c&e panels). 87

Figure 3-17 Fo-Fc electron density map contoured at +3 sigma around the TbXo4 molecule in 10 mM crystallophore HEWL structure. 88

Figure 3-18 Structure of HEWL co-crystallised with 10 mM TbXo4-OH. a) anomalous Fourier map (contour: 5 sigma). 2Fo-Fc electron density map (contour: 1 sigma) of b) TbXo4-OH binding residues and c&e) the TbXo4-OH molecule (two views). d&f) Interaction sites of TbXo4-OH (Two same views as c&e panels). 92

Figure 3-19 Structure of HEWL co-crystallised with 10 mM TbXo4- SO3. a) anomalous Fourier map (contour: 5 sigma). 2Fo-Fc electron density map (contour: 1 sigma) of b) TbXo4- SO3 binding residues and c&e) the TbXo4- SO3 molecule (two views). d&f) Interaction sites of TbXo4- SO3 (Two same views as c&e panels). 93

Figure 3-20 Structure of HEWL co-crystallised with 10 mM TbXo4-NMet2. a) anomalous Fourier map (contour: 5 sigma). 2Fo-Fc electron density map (contour: 1 sigma) of b) TbXo4-NMet2 binding residues and c&e) the TbXo4-NMet2 molecule (two views). d&f) Interaction sites of TbXo4-NMet2 (Two same views as c&e panels). 94

Figure 3-21 Superimposition of HEWL structures with TbXo4 and TbXo4-OH (TbXo4 in green and TbXo4-OH in blue). 95

Figure 4-1 a) Native approach by Tristan Wagner where proteins are isolated from their native host to decipher metabolic secrets (Vögeli et al., 2018). b) SDS page gel of the fractions obtained by native purifications from Methanothermococcus thermolithotrophicus and obtained crystals of proteins from these fractions in presence of 10 mM TbXo4. Native MthAdkA was present in Fraction B (Engilberge et al., 2019). 101

Figure 4-2 Structure of AdkA co-crystallized with 10 mM TbXo4 and soaked in 50 mM solution of TbXo4 (PDB ID 6HF7) a) View of the MthAdkA trimer showing TbXo4 binding asymmetry. b-c) Interaction sites of TbXo4. 102

Figure 4-3 Crystallization diagram of 10mg/ml of MthAdkA WT in the conventional crystallization condition used (Engilberge et al., 2019): PEP629, 100 mM Tris pH= 8.5, 50 mM $MgCl_2$. ○ represents condition with clear drop crystals, ● represents condition with single crystal, * represents microcrystals or/and crystalline precipitates and ˣ represents precipitation. 103

Figure 4-4 Per-residue normalized contact probability ρ of TbXo4 (i.e. the time spanned by the ligand interacting within residue divided by the total simulation time) for MthAdkA, starting from the position from crystallographic structure (orange, 2 TbXo4 in the simulation box) or random position (blue, 10 TbXo4 in the simulation box). The crystallographic sites E29, E50 and D90 are identified with vertical red coloured dashed lines. 104

Figure 4-5 Results of the automated crystallization screening of MthAdkA WT and mutants from Serie 1 (M1 & M2) in the absence/presence of 10 mM TbXo4 (observation after 30 days): unique native hits in blue, common hits (native and TbXo4) in orange, unique TbXo4 hits in grey. 107

Figure 4-6 Crystallization diagrams of 10mg/ml of MthAdkA (a) M1 (E50A) and (b) M2 (D90A) in the conventional crystallization condition (Engilberge et al., 2019): PEP629, 100 mM Tris pH= 8.5, 50 mM $MgCl_2$. ○ represents condition with clear drop crystals, ● represents condition with single crystal, * represents microcrystals or/and crystalline precipitates and ˣ represents precipitation. 108

Figure 4-7 Results of the automated crystallization screening of MthAdkA WT and its mutants from Serie 2 (M3, M4, M5 & M6) in the absence/presence of 10 mM TbXo4 (observation after 30 days): unique native hits in blue, common hits (native and TbXo4) in orange, TbXo4 unique TbXo4 hits in grey.

.....109

Figure 4-8 Results of the automated crystallization screening of TbXo4 with MthAdkA WT and its mutants after 30 days: unique native hits in blue, common hits (native and TbXo4) in orange, TbXo4 unique hits in grey. In a red square marked the second mutation Serie 2. To facilitate analysis, results of Serie 1 are also indicated.....111

Figure 4-9 Crystallization diagrams of MthAdkA (a) M3 (E54Q), (b) M4 (E54D) and (c) M5 (E136Q) in the conventional crystallization condition used in previous work: PEP629, 100 mM Tris pH= 8.5, 50 mM $MgCl_2$. ○ represents condition with clear drop crystals, ● represents condition with single crystal and * represents microcrystals or/and crystallin precipitates. 113

Figure 4-10 Illustration of the influence of a strong binding site on the anomalous Fourier signal to detect minor sites. The red dotted line indicates the 4σ threshold used for peak detection. 114

Figure 4-11 Crystal packing of MthAdkA WT-1 (Table 4-2) and a) the anomalous Fourier map in red representing Tb atom positions b) Residues that interacts with TbXo4 on the surface of the protein; on top the anomalous Fourier map with the residues that interacts with TbXo4 and on the bottom the residues. 119

Figure 4-12 Multiple sequence alignment by Clustal Omega of Lysozymes from Hen Egg-White (LYSC_CHICK), Turkey Egg-White (LYSC_MELGA), Quail Egg-White (LYSC_COTJA) and Peacock Egg-White (LYSC_PAVCR). 121

Figure 4-13 Results of the automated crystallization screening of EWL with TbXo4 after 30 days: number of hits for native protein in blue, common hits (native and TbXo4) in orange, TbXo4 unique hits in grey. 122

Figure 4-14 Crystallization diagrams of (a) HEWL and (b) TEWL after 30 days with different TbXo4 concentrations (0, 2.5, 10, 25 and 100 mM). ○ represents clear drops, ● represents condition with single crystal, * represents microcrystals or/and crystalline precipitates and ˣ represents precipitation. 125

Figure 4-15 Crystallization diagrams of (a) QEWL and (b) PEWL after 30 days with different TbXo4 concentrations (0, 2.5, 10, 25 and 100 mM). ○ represents clear drops, ● represents condition with single crystal, * represents microcrystals or/and crystalline precipitates and ˣ represents precipitation. 127

Figure 4-16 Structure of TEWL co-crystallised with 10 mM TbXo4. a) anomalous Fourier map (contoured at 5 sigma) b) 2Fo-Fc electron density map superimposed to interaction sites of Tb (contour: 1 sigma). Tb atoms are represented as cyan sphere. 131

Figure 4-17 Structure of QEWL co-crystallised with 10 mM TbXo4. a) anomalous Fourier map (contour: 5 sigma) b) 2Fo-Fc electron density map superimposed to TbXo4 molecule (contour: 1 sigma) c)) 2Fo-Fc electron density map corresponding to interaction sites of TbXo4 (1 sigma) d) Interaction sites of TbXo4. e-f) Superposition of TbXo4 in QEWL structure (TACN in green) and HEWL structure (palegreen) co-crystallized with 10 mM TbXo4 (TACN in lightblue). Tb atoms are represented as cyan sphere. .133

Figure 4-18 Structure of PEWL co-crystallised with 10 mM TbXo4. a) anomalous Fourier map superimposed to Tb atoms (contour: 5 sigma) b) 2Fo-Fc electron density map superimposed to TbXo4 molecule (1 sigma) c)) 2Fo-Fc electron density map corresponding to interaction sites of TbXo4 (1 sigma) d) Interaction sites of TbXo4. e-f) Superposition of TbXo4 in PEWL structure (protein in grey, symmetry in wheat; TACN in green) and HEWL structure (pink, PDB ID 6F2I) in the presence of 100 mM TbXo4 site 2 (TACN in lightblue). Tb atoms are represented as cyan sphere. 135

Figure 4-19 Multiple sequence alignment by Clustal Omega of Lysozymes from Hen Egg-White (LYSC_CHICK), Turkey Egg-White (LYSC_MELGA), Quail Egg-White (LYSC_COTJA), Peacock Egg-White (LYSC_PAVCR) Ostrich Egg-White (STRCA) and Softshell Turtle Egg-White (LYSC_PELSI)136

Figure 4-20 Interaction rates of TbXo4 with EWL residues during MM/QM molecular dynamics simulation of 5 times 200ns duration.....138

List of Tables

Table 1-1 Experimental phasing statistics obtained for the eight proteins used . Different de novo phasing methods, using the strong anomalous signal of TbXo4, were evaluated (Engilberge et al., 2017) 41

Table 2-1: List of proteins, Vectors and Antibiotic Resistance (AR), Species of origin, expression strain, correlation coefficient(ε) and their theorical molecular weight. 51

Table 2-2: Screening kits used during this thesis. 57

Table 3-1 Crystallization conditions of HEWL with Xo4 variants. 83

Table 3-2 Structural data collection and refinement statistics of HEWL in the presence TbXo4. 85

Table 3-3 Structural data collection and refinement statistics of HEWL in the presence Xo4 variants. 89

Table 3-4 Occupation rate of TbXo4 variants in HEWL during the molecular simulation. 96

Table 4-1 List of MthAdkA mutants and the corresponding mutations 105

Table 4-2 Crystallization conditions of MthAdkA WT and mutants and anomalous Fourier signal positions (*signal between 4 and 6 σ). 117

Table 4-3 EWL naturel mutations compared to HEWL 121

Table 4-5 Structural data collection and refinement statistics of HEWL and TEWL in the presence of 10 mM TbXo4. 129

Table 4-6 Structural data collection and refinement statistics of QEWL and PEWL in the presence of 10 mM TbXo4. 130

List of Abbreviations

CNT : Classical Nucleation Theory
CCP4 : Collaborative Computational Project No. 4
DFT : Density Functional Theory
DOTA : 1,4,7,10-tetraazacyclododecane- 1,4,7,10-tetra-acetic acid
DO3A : 1,4,7,10-tetraazacyclododecane- 1,4,7-tri-acetic acid
DTPA : Di-ethylen tri-amine penta-acetic acid
DTPA-BMA : diethylenetriamine pentaacetate-bis(methylamide)
EMBL : European Molecular Biology Laboratory
ESRF : European Synchrotron Radiation Facility
HEPES : Acide 4-(2-hydroxyethyl)-1-piperazine ethane sulfonique
HEWL : Hen Egg-White Lysozyme
HPDO3A : 10-(2-hydroxypropyl)-1,4,7-tetraazacyclododecane-1,4,7-triacetic acid
HTX-lab : High Throughput Crystallization Laboratory
IBS : Institut de Biologie Structurale
IPTG : Isopropyl β-D-thiogalactopyranoside
MIP : Molecularly Imprinted Polymers
MMS : Microseed Matrix Screening
MPD : 2- Methyl-2,4-pentanediol
MR : Molecular Replacement
MX : Macromolecular Crystallography
NMR : Nuclear Magnetic Resonance
OD : Optic density
OEWL : Ostrich Egg-White Lysozyme
PDB : Protein Data Bank
PEG : Polyethylen Glycol
PEWL : Peacock Egg-white Lysozyme
PSB : Partnership for Structural Biology
QEWL : Quail Egg-White Lysozyme
RMSD : Root Mean Square Deviation
SDS : Dodecylsulfate de sodium
SER : Surface Entropy Reduction
SOLEIL : Source Optimise de Lumière d'énergie Intermédiaire du LURE
STEWL : Softshell Turtle Egg-White Lysozyme
TACN : Triazacyclononane
TdTau : Thaumatin
TEW: Tellurium-Centered Anderson–Evans Polyoxotungstate
TEWL : Turkey Egg-White Lysozyme
TRIS : 2-amino-2-hydroxym.thyl-1,3-propanediol
Xo4 : Crystallophore

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

Structural biology is a field that seeks to understand the secrets of biological macromolecules and examine their structures and conformations in order to decipher their interactions, functions and roles that they play in biological processes or in diseases. This eventually has opened up new avenues for medical research.

Structural biology involves a variety of techniques that permits to determine macromolecules structures at the atomic scale, with Nuclear Magnetic Resonance (NMR), Electron Microscopy (Cryo-EM) and X-ray Crystallography are the most important techniques.

Nowadays, Macromolecular crystallography (MX) is challenged by Cryo-EM for structure determination (Subramaniam, 2019), as the later has undergone lots of developments making it an efficient structure determination alternative to X-ray Crystallography.

Despite developments in other techniques, X-ray crystallography remains the approach of choice for obtaining atomic-scale structural data of biological macromolecules, as shown by the chart of the number of released structures per Year for each method in the RCSB Protein Data Bank (PDB) (Figure 1-1). Indeed, MX is still considered the most productive and powerful method to obtain high-precision structural information about biological macromolecules, as 86.17% of structures deposited in PDB were determined using this method. This rate is related to (1) the automation of the crystallization process (1.4) by using smaller volumes of biological material and by screening several physico-chemical conditions through numerous crystallization kits available in the market (Brown et al., 2003; Santarsiero et al., 2002), (2) X-ray source development such as tunable micro/nano-focus synchrotron beamlines, their automation as well as new sources with increased capabilities such as X-ray free-electron laser (XFEL) (Emma et al., 2010), (3) Detector developments that have enabled fast data readout, elimination of charge sharing noise, improvement of pixel size and energy filtering (Hülsen et al., 2006). These improvements have provided the groundwork for further advances, for example, in time-resolved crystallography as illustrated by the work on homo-dimeric fluoroacetate dehalogenase (FAcD) from *Rhodopseudomonas palustris*, highlighting its allosteric regulation (Mehrabi et al., 2019) or in developments in drug discovery using fragment

screening approaches (Blundell et al., 2002; Davies & Tickle, 2011; Larsson et al., 2011).

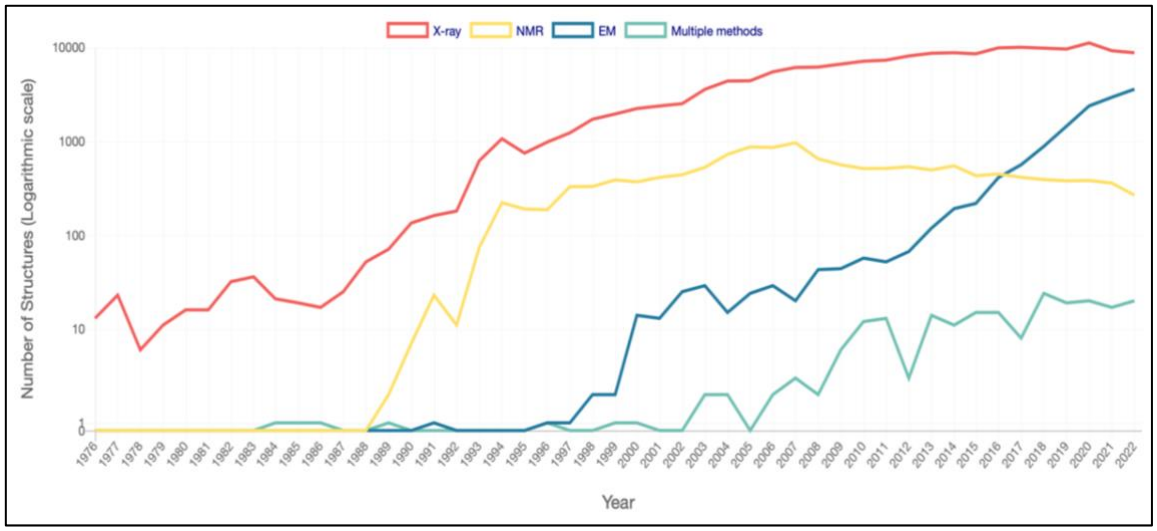


Figure 1-1 *Number of released PDB structures per Year for each method (RSCB PDB Nov-2022)*

The flowchart showing the experimental steps performed during structure determination by X-ray crystallography is presented in Figure 1-2.

Before crystallization, a sufficient amount of the protein target must be obtained first, either directly from its natural source or by overexpressing the target protein gene in an expression system such as the commonly used one *Escherichia coli (E. coli),* then the protein is extracted and purified. Thereafter, a crystal of the protein target is obtained from the crystallization and optimization process. Then, from the synchrotron radiation source, the diffraction data are collected in form of diffraction patterns which later on will be processed in order to calculate the electron density map corresponding to the protein of interest. Finally, a model of its structure is proposed by interpreting the electron density map and further improved and refined.

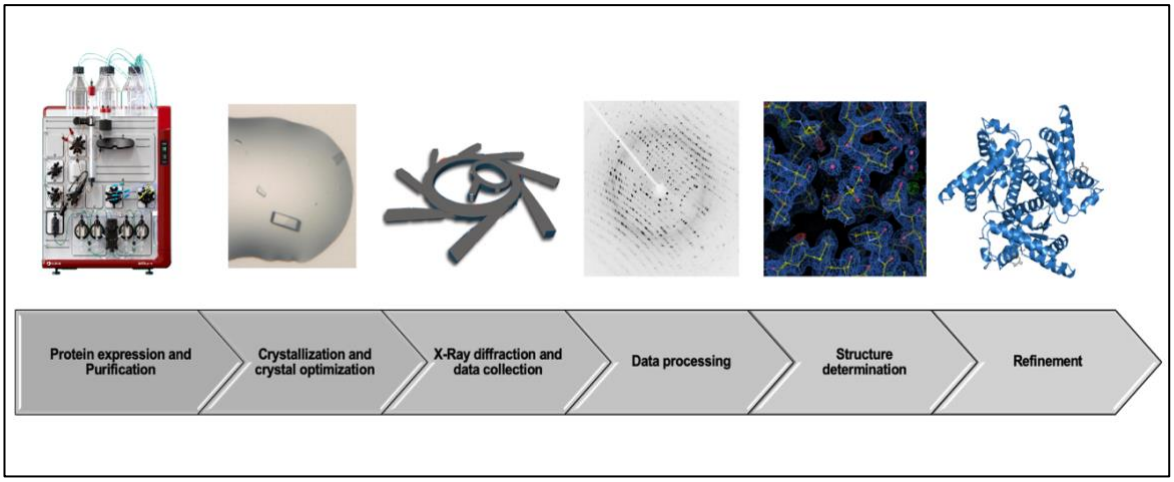


Figure 1-2 *X-Ray crystallography workflow for the determination of protein structure.*

According to the statistical analysis made by Structural Genomics Consortium, only 10% of the purified proteins have their structure determined by X-Ray crystallography (Figure 1-3). Indeed, despite all the development in MX mentioned previously, crystallography still suffers from three main issues:

1) Protein expression and purification: According to the statistics related to the structure determination process from 2014, only 38% of cloned proteins were expressed, of which about half could be purified.
2) Production of high quality well-ordered crystals to ensure the X-ray diffraction experiments. Only roughly 20% of purified proteins led to crystal making crystal production the main bottleneck.
3) Solving the phase problem. Indeed the detector used during the diffraction experiments gives us diffraction patterns allowing us to only obtain the intensities of the diffracted spots, but it does not give us the corresponding phase information. This is the phase problem inherent to crystallography. Noted that several techniques have been used to address this issue, such as: a) Molecular Replacement (MR) by using the phases calculated from models of proteins structurally related to our target, or b) by introducing a sub-structure (heavy atoms) that can be fully resolved thanks to the experiment. The sub-structure can be introduced by amino-acid substitution of typically methionine by seleno-methionine (Doublié, 2007) or by co-crystallization or soaking of the crystals in heavy atom solutions. The initial set of phases can be derived thanks

to the difference introduced by the substructure through either Single/Multiple Isomorphous Replacement (SIR, MIR) techniques or approaches based on Anomalous Scattering at Single or Multiple Wavelengths (SAD, MAD), or by the combination of the two techniques through Single or Multiple Isomorphous Replacement with Anomalous Scattering (SIRAS, MIRAS) (Hendrickson, 2014; Taylor, 2010). In addition to the existing techniques, we would like to note that with the development of artificial intelligence, the phasing problem is less an issue, as programs such as Alphafold (Jumper et al., 2021; Varadi et al., 2022) and RoseTTAFold (Baek et al., 2021a) were developed (Barbarin-Bocahu & Graille, 2022). Indeed, these two programs are able to predict a protein's 3D model from its amino acid sequence. Such model can then be exploited for MR.

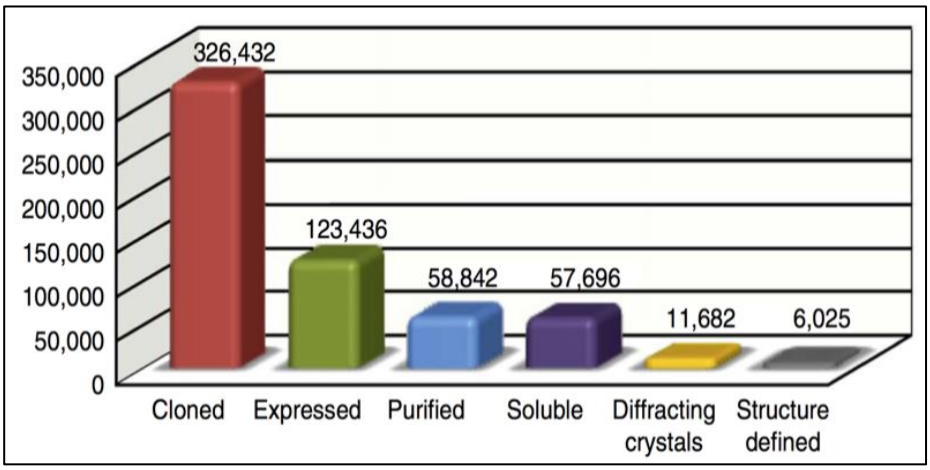


Figure 1-3 *Statistics related to the structure determination process. From (Khurshid et al., 2014)*

In the following part of the introduction, I will first present the considerations to be taken into account to get good quality samples for crystallization. Then, I will detail the theoretical aspects of crystallization process as well as the parameters influencing it and the innovations made in the last two decades addressing the problem of protein nucleation. I will end with the objectives of my thesis in relation with the crystallization issue.

1.1 Getting a quality protein sample

Obtaining a well purified and homogeneous biological macromolecule is considered a crucial and mostly challenging step for the determination of three-dimensional structures by crystallography (Giegé et al., 1986). Other than its purity, protein's quality should be assessed by its integrity, homogeneity, stability as well as its activity over time (Raynal et al., 2014). During this thesis, we have tried to control, as far as possible, the parameters mentioned above for all the model proteins purified and studied, to limit bias in the crystallization outputs.

In this paragraph, I will thus mention the methods used during this thesis in order to obtain protein samples of sufficient quality for crystallization. The protein used in this thesis are either commercial proteins that were used without any further purification steps, or recombinant proteins whose genes have been expressed in *E. coli* or proteins purified natively from the organism. The list and the purification protocols of the proteins used during this thesis are detailed in the Materials and Methods section. After different steps of purification in order to remove the contaminating proteins, all the purified proteins during this thesis were deposited on a gel filtration column as a final step of purification, to get the sample in their respective stock buffer and to assess the purity and the mono-dispersity of our samples.

Following each purification steps, we usually performed SDS-PAGE electrophoresis analysis that allows us to verify the presence of the protein of interest. This technique allows the separation of proteins according to their molecular weight, after denaturing the non-covalent oligomeric states of the proteins. This technique also allows the tracking of the purity of the sample during purification processes.

Then, before exploiting them for screening of the crystallization conditions, the samples were sent to the IBS mass spectrometry platform to verify that the purified protein had the expected sequence without truncation or degradation during expression or purification.

In addition, it has been shown that thermal stability of the sample protein, measured by the half denaturation temperature (Tm), is a critical parameter and may estimate the crystallizability of the protein (Dupeux et al., 2011; Ericsson et al., 2006). For the proteins that have an Tm lower than 45°C, it is recommended to perform the crystallization screening at 4°C. However, for protein with a Tm higher than 45°C, screening at 20°C is advisable (Dupeux et al., 2011; Szilágyi & Závodszky, 2000).

Therefore, the proteins used during this thesis have half denaturation temperatures (Tm) above 65°C, hence the crystallization screening of these proteins were performed at 20°C.

Despite all the steps to characterise the quality of the biological macromolecules and even with 100% purity samples, biological macromolecules remain resistant to crystallization as each molecule can present several conformations in solution (Parak & Frauenfelder, 1993), which may make it difficult to obtain well-ordered crystals.

1.2 Crystallization theory

1.2.1 Phase diagram

Crystallization is a process of forming a crystalline solid form from a liquid solution by phase transition phenomenon called nucleation. This phase transition is formed by the interaction of the protein with its surrounding leading first to the formation of a partially ordered and relatively small aggregates called nuclei (McPherson & Gavira, 2014).

The nuclei will be the origin of the formation of a periodically well-ordered crystal, through a phenomenon called crystal growth, by successive addition of macromolecules leading to larger crystals.

For biological macromolecules, the stochastic and empirical nature of the primary nucleation process, due to spontaneous formation of nuclei, is still globally not understood. Hence, understanding the thermodynamics of crystallization is critical to correctly determine the crystal structure. Therefore, a phase diagram can be defined as a graphical representation of the thermodynamic behaviour of a system. It is also an important tool to understand the influence of several relevant variables on the protein solubility. Indeed, the phase diagram is a graphical representation that depicts the various states of a material (i.e. solid or liquid), as a function of a parameter such as temperature, pressure and chemical characteristics of the crystallization solution i.e precipitant concentration, pH, buffer composition, ionic strength, and other additives.

In Figure 1-4, we present most common form of phase diagram for proteins which is a two-dimensional phase diagram (E. E. G. Saridakis et al., 1994) representing the behaviour of the protein as a function of its concentration and another parameter, usually the concentration of precipitant/crystallising agent (Figure 1-4), with all other

parameters considered as constant. We would like to mention that it is also possible to held a 3D plot phase diagram with the variation of three different parameters (Sauter et al., 1999).

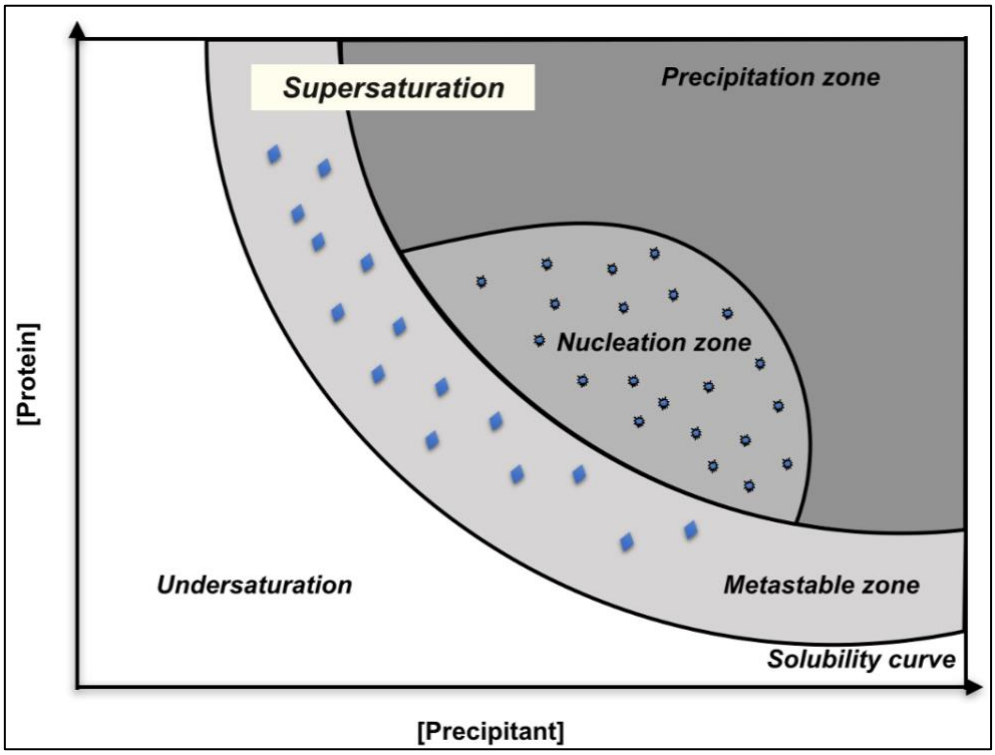


Figure 1-4 *Protein crystallization phase diagram based on two variables, protein and precipitant concentrations.*

From the phase diagram Figure 1-4 we can distinguish four areas:

1) Undersaturation: represented by the area in white where the protein is completely dissolved and neither crystal nucleation nor growth can occur. This area is separated from the supersaturated area by the solubility curve, depicted as the maximum amount of biological macromolecules that can be dissolved in a given volume of solvent.

The next three zones are included in the supersaturation zone, where the protein is in excess in solution.

2) Metastable zone: favouring the crystal growth with no nucleation that will occur.
3) Nucleation zone: this third area corresponds to spontaneous nucleation with the formation of nuclei of proteins. However in this area crystal growth is less favoured compared to the metastable zone. Thus it is important to be in this

zone first, in order to initiate the formation of the crystal and then go through the metastable zone for the crystal growth.

4) Precipitation: High supersaturation area characterized by high nucleation rate leading to excessive attractive protein-protein interactions, therefore, protein precipitation.

We would like to note that the phase diagram presented in figure 1-4 is related to the variation of the concentration of the protein in function of the concentration of the precipitates. Hence, the phase diagram of a given protein is specific to each varied parameter, such as the temperature, pH, additives, etc…

Moreover, we may consider the crystallization process as two stages comprising nucleation and crystal growth. Indeed, both steps depend certainly on the supersaturation state for them to occur.

Therefore, in order to generate the supersaturated state of the solution, the protein should be present in the solution in excess of its solubility limit. Once the saturation limit of the protein in solution is reached the formation of a crystal may occur.

For that, two parameters should be modulated to obtained this state. First, the crystallization cocktail has the ability to reduce the protein solubilization by disturbing the interactions between the solvent and the protein. The optimization of these parameters will be presented in part 1.4. Second, the change of proteins' properties by protein surface mutations using SERp (Surface Entropy Reduction prediction) approach. This approach consists of identifying the most suitable sites for mutation in order to enhance the crystallizability of the protein, which may lead to the reduction of protein solubility, leading thereafter to an increase of the attraction between their surface. This approach will be presented in chapter 4.

1.2.2 Nucleation thermodynamic

As many thermodynamic factors are involved in crystal growth, understanding these factors is critical for successful growth. One of these parameters is characterized by the force driving the protein to come out of its solution and crystallise.

Thermodynamically, this driving force is represented by supersaturation.

Theoretically, the driving force for protein crystallization is represented by Δμ which is the difference in chemical potentials of a protein in solution at its current state (μ_L) and in solution's equilibrium state (μ_S) (Boistelle & Astier, 1988):

$$\Delta\mu = \mu_L - \mu_s = kT.ln\frac{C}{C_s} \quad (1)$$

With k : Boltzmann's constant (1.38 x 10^{-23} $J.K^{-1}$)

T : Temperature (K)

C : Initial concentration of the protein in solution before crystallization.

C_S : Concentration of the protein in the solution's equilibrium state (solubility).

By calculating this difference in chemical potentials, we can determine if the protein has a tendency to change its phase from a liquid to a solid state, reaching the super saturation state. Using the ratio of $\frac{C}{C_s}$ referred as the degree of super-saturation (β), and σ (β-1), both parameters allow to evaluate the supersaturation.

$$\beta = \frac{C}{C_s} \qquad or \qquad \sigma = \beta - 1 = \frac{C - C_s}{C_s} \quad (2)$$

if $\beta < 1$, $\ln \beta = \frac{\Delta\mu}{kT} \sim \sigma$, the solution is considered undersaturated while if $\beta > 1$ the solution is supersaturated.

Moreover, as the nucleation is the most important step during crystallization, it involves a certain amount of energy. This energy is addressed as the Gibbs free energy (ΔG) also known as the free activation germination energy. This energy (ΔG) related to the activation of the formation of nucleus, described below (eq 3-7; Figure 1-5), comes from the thermodynamics of homogeneous primary nucleation (nucleation without impurity), referred to the classical nucleation theory (CNT)(Becker & Döring, 1935; Volmer & Weber, 1926). Moreover, for the nucleation to occur, the driving force (Δμ) representing the supersaturation state should be constant with the variation of the solution volume energy (ΔG_v), which depends of the number of molecules in the cluster (n; eq.4) as the formation of the cluster decreases the surface free energy (ΔG_s) of the cluster (eq. 3).

$$\Delta G = \Delta G_v + \Delta G_s \quad (3)$$

$$\Delta G = -n\Delta\mu + S\gamma = -nkT(\ln\beta) + S\gamma \quad (4)$$

With n: number of molecules that forms the nuclei

k: Boltzmann's constant (1.38 x 10^{-23} $J.K^{-1}$)

T: Temperature (K)

β: Super-saturation degree

S: Area of the nuclei (m^2)

γ: Interfacial tension or surface free energy between the cluster and the solvent.

Assuming that the cluster is spherical with radius r and molecule's volume (Ω), the previous equation will be:

$$\Delta G = -\frac{4\pi r^3}{3\Omega} kT(\ln\beta) + 4\pi r^2\gamma \quad (5)$$

The influence of the volume and surface energy on the free energy function is shown by the evolution of ΔG as a function of the radius of the nucleus (r) (Figure 1-5).
From a certain critical size of a nucleus (r*) (eq 6), the ΔG function passes through a critical value ΔG * (eq 7). At this critical value (r^*), the nucleus is in an unstable equilibrium with the solution. If a molecule is removed r < r*, the system will always seek to minimize its energy, therefore the nucleus will dissolve. However, If a molecule is added r > r*, the nucleus will be stable allowing for its growth.

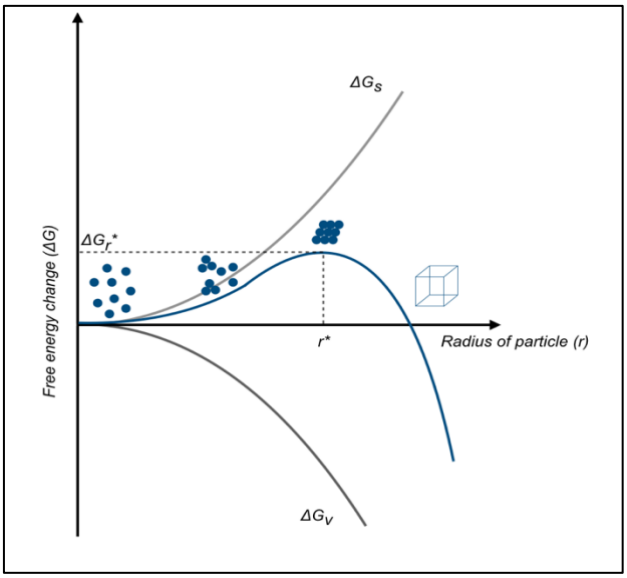


Figure 1-5 *The variation of the Gibbs Free energy (ΔG) according to particle radius (r) during nucleation phenomenon.*

By examining equation (6), the critical nucleus r* shows that it is dependent of the interfacial surface energy (γ), the temperature and the supersaturation degree (β). Hence, r* is smaller as the temperature and the supersaturation are greater, and as the crystal-solution interfacial energy is smaller.

$$r^* = \frac{2\Omega\gamma}{kTln\beta} \qquad \textit{Gibbs-Thomson-Freundlich} \ (6)$$

As for ΔG^*, equation (7) shows that the energy required to form the nucleus depends proportionally on the interfacial surface energy (γ) and inversely proportional to the temperature and the supersaturation degree (β). Hence, the more the solution is concentrated, therefore supersaturated ($\beta > 1$), the lower the free energy ΔG^* is, thus the easier the nucleation is achievable. Same for the temperature.

$$\Delta G^* = \frac{16\pi\Omega^2\gamma^3}{3(kTln\beta)^2} = \frac{1}{3}\,(4\pi r^{*2}\gamma). \qquad (7)$$

Moreover, equation (7) also shows that the energy required to reach a stable nucleating point is equal to one third of the energy required to create its surface (ΔG_s) and that the smaller r* is, the smaller ΔG^* is.

For the interfacial surface energy (γ) parameter, that corresponds to the cluster-solution interactions, this parameter depends, to some extent, on the affinity and the solubility that the solvent has for the macromolecule i.e. the higher the affinity, the greater the solubility with the decrease of γ, thus the easier is the nucleation because there are more encounters between monomers in solution (Boistelle & Astier, 1988). Therefore, as the solution compositions acts by promoting the protein adsorption on the cluster surface, the interfacial energy γ can only decrease when the adsorption of proteins increases, thus increasing the solubility making the nucleation easier to occur.

1.3 Crystallization methods

Currently, there are several different methods used for the crystallization of biological macromolecules that allow to generate crystals from saturated solutions. These methods are detailed in the following paragraphs, with their respective paths are indicated in the phase diagram in Figure 1.6.

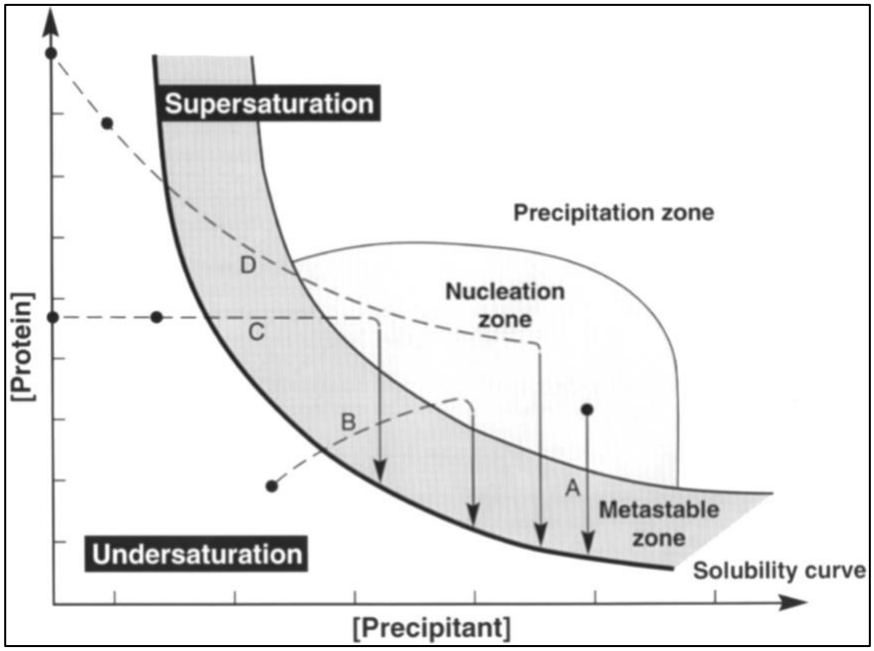


Figure 1-6 *Protein crystallization phase diagram based on two variable parameters, protein and precipitant concentrations (N. E. Chayen, 1998) in four major crystallization methods: A. Batch, B. Vapour Diffusion, C. Dialysis and D. Free-interface diffusion.*

Experimental set-up used to exploit these crystallization methods are summarized in Figure 1.7.

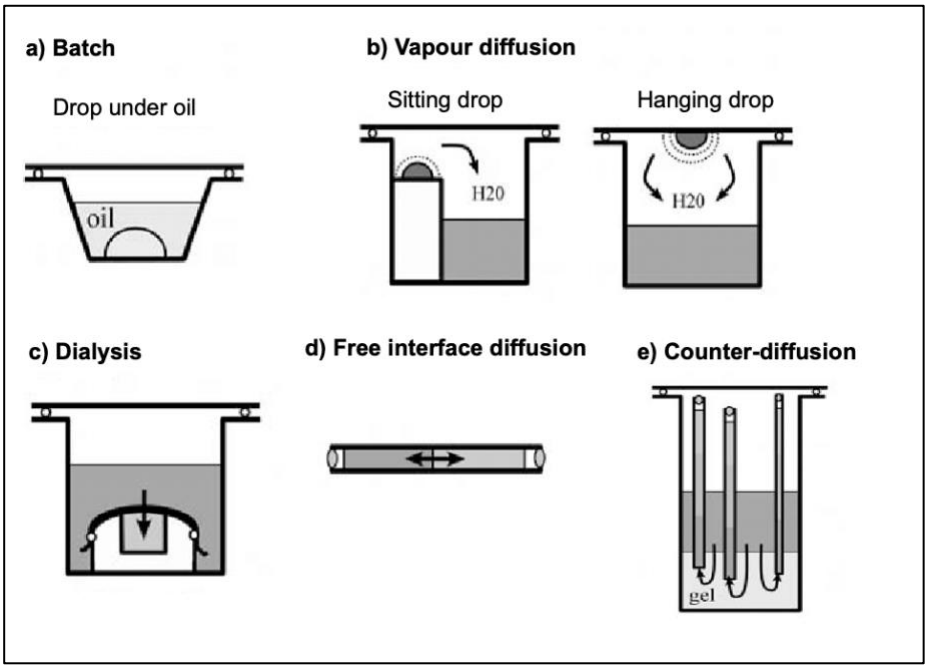


Figure 1-7 *Principles of major methods used to crystallize biological macromolecules (a) a version of batch crystallization. (b) Vapour-diffusion crystallization with sitting and hanging drops. (c) Dialysis method with Cambridge button. (d) Crystallization by free-interface diffusion in a capillary where two solutions of equivalent volume are brought into contact. (e) Counter-diffusion setup in X-ray capillary tubes pinched in a gel layer. (adapted from (Sauter et al., 2012)).*

1.3.1 Batch

The batch method is a micro/nano scale batch crystallization method where the protein is directly mixed with the crystallising agents that aim to obtain the supersaturation at once after mixing, inducing often an excessive nucleation. The process of crystallization for this method is represented in Figure 1-6 A. In the technique named micro-batch under oil crystallization, a derived technique of batch crystallization, 1-2 µl of the protein-crystallization cocktail mix is incubated under a layer of paraffin oil (or silicone oil), which is thick enough to prevent the evaporation of small volumes during the crystallization process (Figure 1-7a) (N. Chayen & al 1990). Noted that a thickness of 4 mm (equivalent to 6 ml) of oil should be use, as a thinner layer may allow the evaporation which may result in drop drying.

The batch method is especially used for optimisation by strictly controlling the crystallization environment and, nowadays, to produce large amount of samples for serial crystallography for example (Stohrer et al., 2021). Indeed, as the conditions in this method are expected to be constant, it is easier to monitor the sample history and obtain reproducibility.

1.3.2 Vapour Diffusion

In contrast to batch method, vapour diffusion methods (Hampel et al., 1968) are dynamic systems in which protein solution is undersaturated at the beginning of the experiment and is progressively driven to supersaturation by equilibration with a reservoir solution which contains the crystallising agents. This method is the most popular method to obtain protein crystals. The process of crystallization in vapour diffusion technique is illustrated in Figure 1-6 B. It is widely used and it includes several techniques such as sitting drop technique where the drop of the mix is placed on a support in the well or hanging drop technique where the crystallisation drop is deposited on a siliconized cover slip (Figure 1-7 b). Both techniques have same principle, a drop of 100nl to 1 µl of protein solution is mixed with an equivalent volume of crystallizing agent solution and equilibrated, in an hermetically sealed well, against 45 µl to 1000 µl of the crystallization solution in the reservoir. At the beginning of the crystallization process, the concentration of the crystallising agent is lower in the drop than in the reservoir. Therefore, the water or organic solvents, as the most volatile agent in the solution, will diffuse from the drop to the reservoir until the equilibrium between the drop and reservoir is achieved. Thus, the decrease in the volume of the drop causes the concentration of the constituents in the drop to rise, including the protein one. This results in a change in the concentration of the protein from an initial concentration C to a final concentration C_f, reaching the supersaturation zone where the nucleation is favoured. When stable nucleation points appear, crystal growth begins with the decrease of the concentration of soluble protein in the drop.

1.3.3 Dialysis

Dialysis is a commonly used method that permits the variation of several parameters that affect protein crystallization (Figure 1-7 c). In this method, the protein is placed in

a dialysis button closed by a semi-permeable membrane which allows the diffusion of small molecules from the crystallization solution. Each dialysis button is placed in a well that contains a precipitant solution. Then the well is hermetically sealed with vacuum grease and glass coverslips. In this method, membrane molecular weight exclusion size, the ratio of the concentrations of the precipitant inside and outside the dialysis chamber and the temperature are main parameters to achieve the equilibrium. Also, the geometry of the dialysis cell may influence the crystallization process.
This method requires rather high amounts of protein (10 µl to 100 µl) depending on the used equipment (Sauter et al., 2012). In Figure 1-6 c, Cambridge button (Reid et al., 1973) is shown. The process of crystallization with dialysis method is illustrated in Figure 1-6 C.

1.3.4 Free-interface and counter-diffusion methods

Free-interface and counter-diffusion methods (García-Ruiz & Moreno, 1994; Salemme, 1972) are two methods that require direct diffusion, of the crystallization agent into protein solution, with minimal convection in the channel due to the small diameter of the capillaries (0.1 mm) (Figure 1-7 d & e).
Free interface diffusion (Salemme, 1972) consists of pouring gently the less dense solution into the densest one to avoid too rapid mixing of the two solutions in the capillaries. Then, the constituent of each solution will diffuse into the other compartment. The process of crystallization with Free interface diffusion method is illustrated in Figure 1-6 D.

Counter-diffusion method (García-Ruiz, 2003; García-Ruiz & Moreno, 1994; Ng et al., 2003; Otálora et al., 2009) consists of incorporating several capillaries filled with protein solution into a gel base formed from agarose or silica in a small container where the concentrated crystallization agent solution is poured over its surface. Then, the precipitants will enter the capillaries from the gel and form an upward gradient in the capillary, promoting crystallization along the way as go up by diffusion.

During the whole of this thesis work, we had used the vapor diffusion method, the most popular method that is used in crystallization experiments. Also, we had access to an automation facility, that allows the screening of hundreds of crystallization conditions,

named high throughput crystallisation platform of EMBL-Grenoble, called HTX-lab. This platform allows to typically screen around 576 crystallisation conditions from six crystallization kits commercially available (Chapter 2, Table 2-2). The screenings were performed for each protein using the sitting drop technique, where a volume of 100 nl of protein is mixed with 100 nl of precipitant solution and this drop is then equilibrated against a well of 45 µl. This crystallization setup has allowed us to determine some successful conditions that have led to crystallization of our proteins of interest.

In addition to the selection of the best crystallization methods to obtain high-quality crystals, the formation of a crystal depends on other parameters which affect proteins' interactions. These parameters such as the quality of the protein, which was mentioned above, precipitants and other chemicals such as protein and precipitant concentrations, pH, and physical parameters such as temperature, pressure should be optimized.
Therefore, in the following paragraph, we will present the factors and the methods, in order to optimize a crystallization experiment.

1.4 Optimization of crystallization

Obtaining a well diffracting crystal of a macromolecule depends on the interactions between the macromolecules in solution as well as the parameters of the growth environment (physicochemical composition) such as precipitant nature and concentration, buffer, pH, temperature, impurities. By carefully controlling these variables, it is possible to produce well diffracting crystals.
Also, obtaining crystals that are not suitable for diffraction (too small, not well diffracting crystal etc…) and when optimizing the quality of the protein, its concentration, sample volume, crystallization methods and crystallization conditions are not sufficient, techniques like seedings are commonly used for crystallization troubleshooting. Moreover, in case of extensive screening with no spontaneous nucleation observed, introducing chemical compounds or natural probes named "nucleating agents" may help to initiate the nucleation, therefore facilitates the crystallization. All these methods will be detailed in the following parts.

1.4.1 Optimization of crystallization conditions (crystallization cocktail)

In order to obtain a well diffracting crystal, several controls and optimization should be done, such as controlling the quality of the protein, presented in part 1.2, and optimizing the crystallization conditions. In this part, I will present in details the influence of the precipitant agents and briefly the influence of pH and temperature on the crystallization.

Precipitating agents affect the interactions between the biological macromolecule and the solvent. In this context, a great variety of distinct precipitating agents exists, such as salts (NaCl, KCl, $MgCl_2$,$CaCl_2$, various ammonium and phosphate salts…), polyethylene glycols of various sizes and organic solvents (methanol, propanol…).

The impact of these precipitants on the macromolecule crystallization depends on their concentration, as their concentration affect the solubility of the macromolecule in solution due to ionic forces. This is represented by the Debye-Hückel theory that describe the relationship between protein solubility (log S) and the ionic strength of the solution (Figure 1-8) using two phenomena “Salting-in” and “Salting-out”.

Salting-in phenomenon refers to an increase in protein solubility with a slight increase in salt concentration, within a low concentration range of salt. This is because salts interact on the surface of the protein and increase its solubility by decreasing the electrostatic energy between the protein molecules.

As for the phenomenon of salting-out, it is the decrease in the solubility of proteins due to the increase in the ionic strength of the solution upon addition of high concentration of salts. This can be explained by the fact that salt molecules compete with protein molecules to bind with water. Thus, protein molecules will tend to associate with each other as protein-protein interactions will become energetically more favourable than protein-solvent interactions.

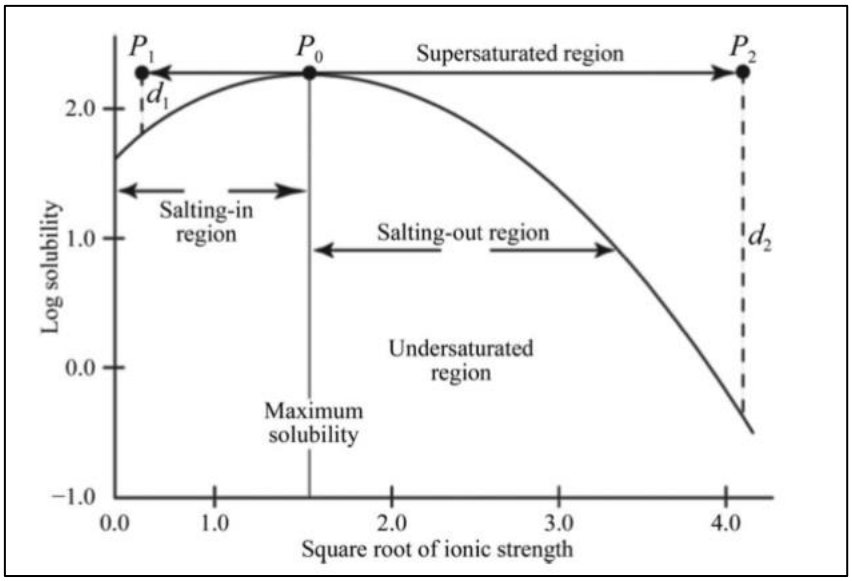


Figure 1-8 *Evolution of the protein solubility (log S) as a function of the square root of ionic strength ($\sqrt{I}$) of the crystallization solution (McPherson & Gavira, 2014).*

Another widely used precipitant, polyethylene glycol, does not present a single conformation in solution. As a consequence, this type of precipitant produces volume exclusion effects and induces the separation of macromolecules from solution (Ingham, 1990; McPherson, 1976). Indeed, as it occupies a large volume of the solvent, it competes with the biological macromolecules for the solvent, causing them to separate, aggregate, and eventually form a solid state (McPherson, 1976).

Moreover, organic molecules (methanol, ethanol, 2-Methyl-2,4-pentanediol (MPD)) tend to reduce the dielectric constant of the solvent and influence the electric field that mediates macromolecular interactions in solution. Also, as the organic solvent concentration increases, the attractive forces between macromolecules increase, decreasing the interfacial surface energy γ between the cluster, thus favouring the solid state (formation of nucleus) (McPherson & Gavira, 2014).

Even though there have not been an identification of a specific type and favourable interaction to the formation of crystalline contacts, studies have shown that hydrophobic interactions are important in order to induce the protein-protein contact in solution, hence induce the nucleation (Nanev et al., 2019; Young et al., 1994).

A key parameter, pH, influences the interactions of macromolecules in solution by having an impact on their solubilities. It is related to its effect on the overall charge of the protein as pH affects the charge of the amino acids present at the protein surface. In addition, the change in temperature can affect pH behaviour of macromolecules as well as the buffer pH, for example TRIS buffer shows a decrease of 0.03 pH units for each increase of 1°C in temperature (Bruins et al., 2007).

Finally, since each protein has a unique length and amino acid sequence, as well as specific conformations, a solubility diagram controlling several parameters associated to the macromolecule and its environment, at the same time, is not experimentally feasible (McPherson & Cudney, 2014).

1.4.2 Seeding

Seeding is one of the simplest methods of controlling supersaturation through nucleation control, promoting low supersaturation and high quality crystal formation. It allows the growth of crystals from a reference crystal of micro or macro seeding type, in the metastable zone where the slow growth of well-ordered crystals is favoured (D'Arcy et al., 2007; Ireton & Stoddard, 2004) .

Similarly, Micro-seeding consists in breaking crystals, formed from previous crystallization trial, directly in the drops by using a probe also named a crystal crusher. Thereafter, the drops that contained the crushed crystals are transferred into a tube which contained a bead, then adding and aspiring 5 or 10 µl of the crystallization reagent from the reservoir to the drop while repeating this process up to 10 times in order to retrieve all the crushed crystals stuck to the slide. This solution is vortexed in order to reduce the remaining crystals into extremely small particles. generating a seeding stock solution. After that, the obtained seeding stock is used for manual or automated Microseed Matrix Screening (MMS) (D'Arcy et al., 2007, 2014). The protocol of MMS consists in either using directly the seeding stock or use a diluted version to control as best as possible the number of crystals. MMS uses this stock to screen again a whole set of crystallization conditions in the presence of the seeds. The seeds can be added automatically to the drop by using crystallization robot, when the capability is available and is the method used nowadays. Historically, manual seeding, through streak seeding, was performed, by using animal hair, bristles or whiskers to add the seed into the drop (E. Saridakis & Chayen, 2009). It should be mentioned that the tools were also directly used to initiate the nucleation due to the

mechanism of action of the hair, which is related to its shape and the keratin present on its surface (E. Saridakis & Chayen, 2009).

Hence, streak seeding was used to generate crystals of good quality even for proteins which are hard to be crystallized such as membrane proteins (MPs) (Kolek et al., 2016; Zhou et al., 2017). Moreover, another technique of seeding, known as cross-seeding can be performed with MPs as well as soluble proteins. This technique consist of using crystals from a protein with similar sequence to the target one in order to do the seeding (Zhou et al., 2017).

1.4.3 Nucleating agents

At the origin, the novelty to produce nucleating agents were due to the fact that some proteins do not produce any crystals with the observation of totally clear drop despite extensive screening. Thus, nucleating agents came to overcome this hurdle and were also used for improving or increasing nucleation. These compounds target the metastable zone in order to avoid the formation of aggregates, therefore promoting crystal growth. There are mainly two types of nucleating agents depending on the way they induce the nucleation. The first one consists in foreign solid compounds or surfaces acting as seeds leading to heterogeneous nucleation. One can mention mineral dusts (McPherson & Shlichta, 1988), natural nucleating additives like horse hair (D'Arcy et al., 2003) and dried seaweed (Thakur et al., 2007), protein thin film templates (Pechkova & Nicolini, 2002), porous materials (Chayen et al., 2001, 2006; Khurshid et al., 2014), and Molecularly Imprinted Polymers (MIPs; semi-liquid) (E. Saridakis et al., 2011). These heterogeneous agents are generally easy to use, stable and do not necessarily require an exact and reproducible quantity between each drop (Khurshid et al., 2014).

As for the second type of nucleating agents, named as homogeneous nucleating, they are nucleating agents that dissolved in the solution of the protein leading to an homogeneous nucleation phenomenon, similar to the results obtained by the precipitates with some proteins. This nucleation type used liquid nucleating agents such as "molecular glue".

1.4.3.1 Porous materials

In 2001, a new approach to protein crystallization has been introduced. This approach consists in using heterogeneous nucleating agents formed from porous material such as silicon (first porous material). Basically, the idea is to use a material with pores' size similar to the target protein dimension (Chayen et al., 2001). Therefore, the pores will favour the formation of ordered nuclei, and thus help crystallisation. Thereafter, other porous materials were proposed as nucleating agents: Bioglass (Chayen et al., 2006), carbon nanotube (Asanithi et al., 2009), nanoporous gold (Kertis et al., 2012; Khurshid et al., 2014)…

In order to induce the nucleation mechanism by the porous material, several properties should be taken in to account. First, the main factor for the success of these materials is the pore size as mentioned previously. The pores size should be close to the size of the targeted macromolecules. Second, the pores will confine the proteins by diffusion and capillarity, causing an increase in macromolecule concentration at the material's surface either due to the mechanism of diffusion or due to weak interactions between the proteins and the surface of the pores, leading to the formation of a nucleus. As the interaction between the pores and nucleus is weak, it is possible for the nucleus to be detached from the pore, thus turned into a seed in the solution.

However, these methods are not suitable for automated crystallization due to the presence of solid materials that should be handled and have demonstrated limited efficiency.

1.4.3.2 Molecularly imprinted polymers (MIPs)

Molecularly imprinted polymers (MIPs), also referred to as "smart materials", are acrylamide-based polymers. The idea is to form these materials in the presence of imprinting proteins close to the size of the protein to be crystallized (E. Saridakis et al., 2011). These imprinting proteins are then extracted after polymerization of the solution, leaving cavities behind them producing a porous material that is supposed to be highly selective and specific of the target proteins. Then, the polymer will be placed in the crystallization drop from which the target proteins will interact specifically with the

polymer in the imprinted regions, favouring the formation of crystal. This technique is quite cheap and easy to use. It is also suited for automated protein crystallization trials.

Despite their compatibility with the crystallization process, all these heterogeneous nucleating agents mentioned above possess disadvantages such as inducing the crystallisation of the components presented in crystallisation media, mainly in conditions with high salt concentration, thus inducing false positive results (Khurshid et al., 2014). In addition, another important disadvantage of these heterogeneous nucleating agents is their incorporation into the screening step. Currently, there is no existing automated system which can add the solid nucleating agents. Only MIPs may be used by crystallization robots as they are semi-liquid.
Therefore, nowadays, a high-throughput crystallization platform is the best option to determine optimal conditions for protein crystallization through evaluation of a large chemical space. However, the macromolecular crystallization process remains a quite expensive and inefficient step with no guarantee of success. Therefore, the necessity to develop new alternatives compatible with automation screening and for which experimental analysis are simpler remains. These alternatives are described hereafter.

1.4.3.3 Silver Bullets

Silver bullets are small molecule cocktails that can be incorporated in the crystallization solution. These molecules promote protein nucleation and growth by stabilizing crystal lattice contacts by their capabilities to generate hydrogen and electrostatic bonding as well as the possibility of modifying the hydrophobic effect (McPherson & Cudney, 2006). These small auxiliary molecules are traditionally called additives and not precipitants as they are used in addition to other precipitating agents. In addition, in order to demonstrate the effectiveness of these small molecules three separate experiments were performed. These experiments evaluated the effects of 200 small molecules, which most of them are considered potential protein ligands, on the crystallization of 81 proteins and viruses by vapor diffusion with only two fixed crystallization conditions at pH 7 (McPherson & Cudney, 2006). The results of these experiments had shown that 85% of the proteins were crystallized, with 54% of these proteins were able to crystallize only in the presence of at least one small molecule.

Hence, the necessity to incorporate one or more small molecules for the crystallization of some proteins. In addition, It has been determined that dicarboxylic acids and diamino compounds of various sizes and geometries are generally the most promising types of small molecules. Moreover, these small molecules induce different conformations of some proteins and shows that it has an effect on the diffraction quality of some crystals (Larson et al., 2007).

Despite their advantages, and in addition to the low probability to crystallize a protein specifically by one of the small molecules, these small molecules may face deleterious effects on the crystal growth due to the masking of their positive contributions by the components present in the crystallization conditions (Larson et al., 2007; McPherson & Cudney, 2006b). This problem may be overcome by adding each small molecule in several crystallization conditions (Larson et al., 2007).

1.4.3.4 Molecular Glues

As mentioned above in part 1.3, obtaining high-quality crystals of a target protein is a major problem related to the nucleation step. Indeed, controlling the nucleation step that governs the crystallization process would address this conceptual problem. Therefore, the search for a 'universal' nucleating agent, referred to a substance that can induce the nucleation of any protein is in progress.

Nucleating agents such as “Molecular Glues” are water-soluble molecules that facilitate crystal formation by stabilizing the contacts between two proteins through supramolecular interactions. They thus present few similarities with silver bullets. These agents are easy to be incorporated directly into the crystallization medium like the precipitants. However, their concentration differs from the concentration of precipitants used for crystallization, as we use lower concentrations. In addition, these water-soluble agents are fully suitable with high-throughput crystallization platforms (Bijelic & Rompel, 2018; McPherson & Cudney, 2006). In this section, we will first present 2 molecular glues: Polyoxometalates (POMs) and Calixarenes, then lanthanide complexes, we develop, that will be detailed in section 1.5.

a) Polyoxometalates (POMs)

Polyoxometalates (POMs) are polynuclear inorganic metal-oxo anions, with metals from transition atoms such as Molybdenum (Mo), Tungsten (W), Vanadium (V), Niobium (Nb) or Tantalum (Ta). These metals are in their highest oxidation states leading to POMs with general formula: $[X_xM_mO_y]^{n-}$ (X: heteroatoms such as P, As, Si and Ge). These compounds have been studied and used in many field such as nanoscience, medicine and most recently macromolecular crystallography (Bijelic & Rompel, 2015, 2017)**.** Initially, POMs were used for phasing, due to the large heavy metal content that they possess (Bijelic & Rompel, 2015, 2018).

Negatively-charged POMs interact mainly with positively charged amino acids present at the surface of the protein via mainly electrostatic interactions (Figure 1-9). They also form hydrogen bonds with proton donating amino acids. Moreover, POMs can interact with hydrophobic amino acids via Van der Waals interactions. Therefore, POM-protein interactions stabilize conformation by stiffening certain particularly flexible regions of proteins and act as a "molecular glue" that stabilizes crystal packing.

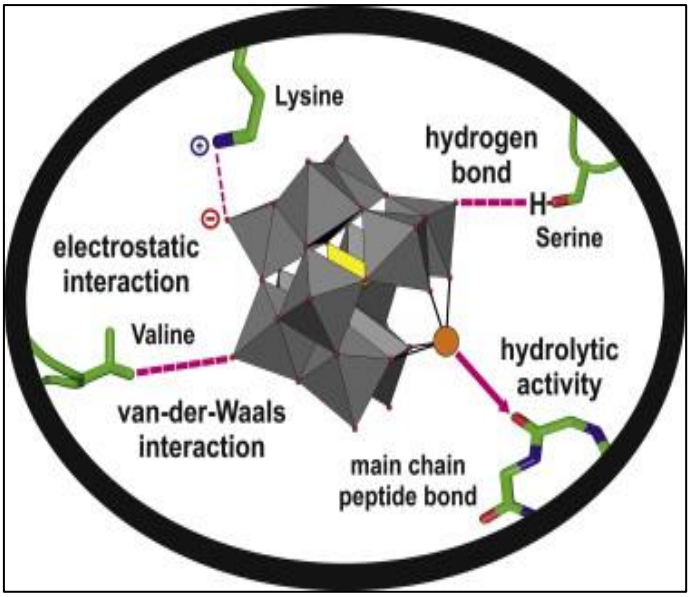


Figure 1-9 *Graphic showing the most frequent POM–protein interactions (Bijelic & Rompel, 2015)*

One of the first polyoxometalate compound that has been referred to as a molecular glue is the Tellurium-Centered Anderson–Evans Polyoxotungstate $[TeW_6O_{24}]^{6-}$ (TEW) (Bijelic & Rompel, 2017) which was used to solve 2 proteins with unknown structures (Mauracher et al., 2014; Molitor et al., 2015) as well as to determine the structure of Hen egg-white lysozyme (HEWL) in a new unknown crystal packing with 4 molecules per asymmetric unit (PDB: 4PHI; Bijelic et al., 2015).

Thereafter, TEW is considered an important crystallization additive as it is very soluble and stable in most of the crystallization solutions, in addition to its ability to interact with the protein without affecting its integrity (Bijelic & Rompel, 2015). It is commercially available as a kit named XP screens. However, as a charge molecules, TEW may contribute to crystallization of some precipitants and thus induce false positive crystallization hits.

b) Calixarenes

Calixarenes (Shinkai et al., 1988) are macrocycles resulting from the reaction of phenols and aldehydes. They are formed of n phenolic units para substituted, the most common being calixarenes with 4, 5, 6, and 8 phenolic units (lower rim; Figure 1-10). Also, several types of calixarenes have been proposed depending on their substitution; like sulphonated (sclxn) (McGovern et al., 2014) and phosphonated (pclxn) (Rennie et al., 2017) calixarenes (upper rim; Figure 1-10). Moreover, calixarene presents a hydrophobic cavity that can contain small molecules or ions.

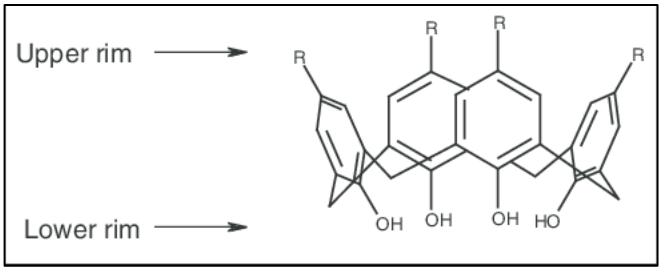


Figure 1-10 *Structure of calix[4]arene. R =* SO_3^-, HPO_3^-, $CH2-HPO_3^-$

These molecules interact by non-covalent bonds with two or more protein surfaces and can promote crystal lattice formation (Alex et al., 2018). The hydrophobic

interactions, with high affinity, is done essentially with lysine (K) or arginine (R) present at the surface of the protein. In the case of sulphonated or phosphonate substitutions, they also form salt bridges, π-π or π-cation interactions with the surrounding amino acids.

However, despite their interesting properties, the study of calixarenes as molecular glue to induce nucleation has been demonstrated on only four small proteins (Cytochrome c , Hen Egg-White Lysozyme (HEWL), *Ralstonia solanacearum* lectin (RSL) and Penicillium antifungal protein (PAF)) (Alex et al., 2018, 2019; McGovern et al., 2012, 2014).

1.5 Way to the third generation of Lanthanide Complexes

1.5.1 Lanthanide complexes : history → TbXo4

Over the past two decades, lanthanide complexes have been proposed by our group initially as phasing agents in order to overcome the phase problem thanks to the significant anomalous contribution of lanthanide ions (Djinovic-Carugo & Carugo, 2015; Hendrickson, 2014; Kahn et al., 1985; Lye et al., 1980). Therefore, three generations of lanthanide complexes have been developed. The first generation that includes a lanthanide (gadolinium) ion chelated by a macrocycle ligand (DOTA, DO3A and HPDO3A) or a polydentate one (DTPA, DTPA-BMA) (Figure 1-11) has been studied during the thesis of Eric GIRARD (1998-2001) and of Meike STELTER (2002-2005) (Girard et al., 2003a; Stelter et al., 2014).

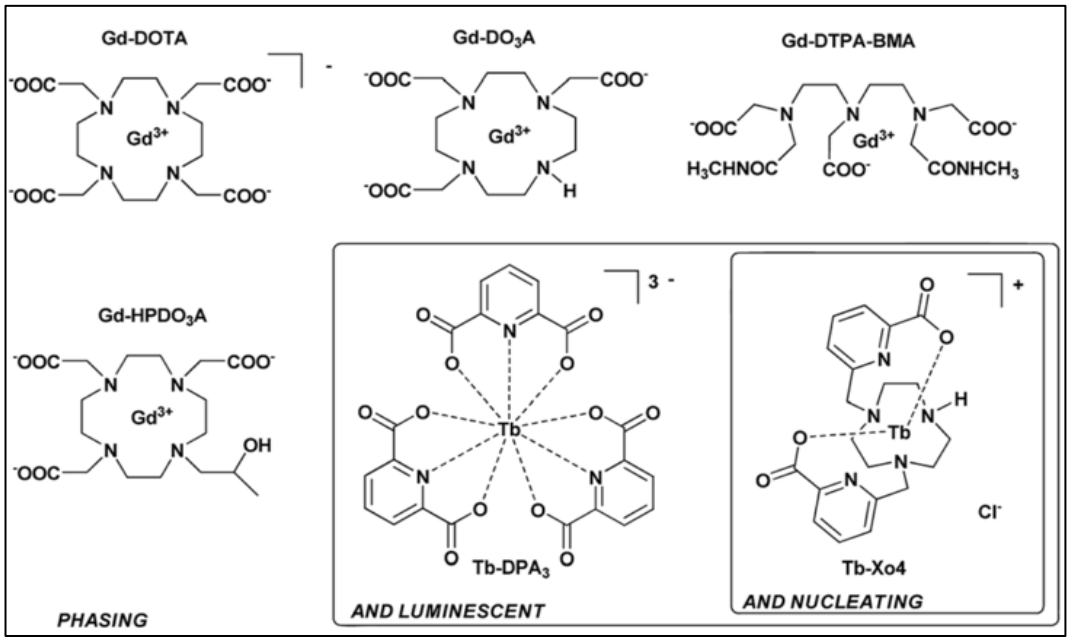


Figure 1-11 *Three generations of lanthanide complexes designed for protein structure determination (Engilberge et al., 2017)*

These complexes have been evaluated on different commercial proteins and have allowed the determination of new proteins structures (Arnoux et al., 2009; Chaudhuri et al., 2003; de Bono et al., 2005; Eichmann et al., 2016; Larivière et al., 2012). Then comes a second generation of lanthanide complexes based on tris-dipicolinate

lanthanide complexes (DPA) (Figure 1-11), that was evaluated by Guillaume POMPIDOR (2004-2007) thanks to a collaboration with Olivier Maury's Team at the ENS of Lyon. During his thesis, he showed that this generation of lanthanide complexes, that possess luminescent properties, can have positive effects on nucleation and crystallization (Pompidor et al., 2008) in addition to being very good phasing agents (Pompidor et al., 2010). Pompidor's thesis was followed by the thesis of Romain TALON (2008 - 2012), dealing with derivatives of DPA complexes. Romain studied the influence of DPA ligand functionalization on crystallization as well as on the phasing using several model proteins (Talon et al., 2011, 2012). However, these DPA-based molecules do not cover wide range of proteins and have shown stability issues in crystallization solutions (such as self-crystallization). Hence the need of a new complex to overcome these drawbacks arises.

Therefore, based on these previous studies and the knowledge accumulated, a third generation of lanthanide-based molecule, later named "Crystallophore" (TbXo4, Figure 1-12), was designed in collaboration with Olivier Maury's Team at the ENS of Lyon.

This cationic complex is formed of a terbium (3+) ion surrounded by a macrocyclic triazacyclononane bis-picolinate ligand and possesses an excellent solubility and stability in the physico-chemical conditions found in crystallization media.

This complex was first studied by Sylvain ENGILBERGE (2014-2017) and had revealed its unique properties, as TbXo4 not only possesses phasing properties, it also promotes crystallization. It is thus an all-in-one tool that overcomes the two main issues of macromolecular crystallography.

In the following paragraphs, I will talk briefly about the phasing property of TbXo4 and I will detail the nucleating properties of TbXo4 as well as our current knowledge on its binding mode.

1.5.2 TbXo4 properties

1.5.2.1 Phasing properties

As mentioned above, lanthanide ions are suitable applicants for anomalous-based phasing of the diffraction data as they exhibit among the largest anomalous contribution (Hendrickson, 2014). In addition, due to their high Z number, they may

be highly efficient for phasing with isomorphous techniques. Therefore, phasing properties of TbXo4 was evaluated by performing anomalous-based experiments on 8 different proteins (Engilberge et al., 2017). For that, diffraction data were collected at the terbium L_{III} absorption edge (wavelength : 1.64895 Å) and SAD or MAD phasing protocols were applied for solving the phase problem (Table 1-1).

Table 1-1 *Experimental phasing statistics obtained for the eight proteins used . Different de novo phasing methods, using the strong anomalous signal of TbXo4, were evaluated (Engilberge et al., 2017)*

Protein	Soaking in 100 mM Tb-Xo4 (soaking time)	Final Tb-Xo4 concentration for phasing (mM)	Phasing method	Resolution (Å)	Space group	FOM after SHARP	% solvent content[a]	FOM after Solomon	Buccaneer[b]	U.A. content (molecules)
HEWL	No	10	SAD	1.8	$P4_32_12$	0.42	41.3	0.88	97.2	1
Thaumatin	No	10	SAD	1.8	$P4_12_12$	0.40	45.6	0.92	75.4	1
Proteinase K	No	10	SAD	1.8	$P4_32_12$	0.31	45.8	0.86	0.0	1
Proteinase K	Yes (120 s)	100	SAD	1.8	$P4_32_12$	0.26	44.5	0.85	86.4	1
Protease1	No	10	SAD	2.0	$P4_12_12$	0.21	61.9	0.95	62.7	3
Protease1	Yes (45 s)	100	SAD	2.0	$P4_12_12$	0.35	60.1	0.96	87.8	3
GRHPR	Yes (120 s)	100	SAD	2.5	$I4_1$	0.18	68.7	0.92	90.8	1
pb9	Yes (45 s)	100	SAD	2.0	$P2_1$	0.58	50.8	0.87	84.6	4
pb6	Yes (300 s)	100	MAD 2λ[c]	2.85	$P2_1$	0.29	65.5	0.79	64.7	4
MDH	No	10	SAD	2.35	$R3$	0.32	50.8	0.94	99.5	4

[a] As evaluated by AutoSHARP. [b] The values indicated correspond to the completeness of chains (%) as provided by Buccaneer. [c] Two-wavelength MAD experiment.

For the determination of the structures, crystals of the eight evaluated proteins were all obtained by co-crystallization with 10 mM TbXo4, the optimal concentration experimentally determined to generate nucleating effects (Engilberge et al., 2017). After data collection, 60 to 100% of protein model were automatically built for the eight proteins (Table 1-1). Some structures of the evaluated proteins were determined after a rapid soaking in 100 mM TbXo4 solution for 45 s to 5 min, in order to reinforce the anomalous contribution.

1.5.2.2 Nucleating Properties

As mentioned, a panel of eight proteins has been initially used to demonstrate the capabilities of TbXo4 (Engilberge et al., 2017) including the nucleating one (Figure 1-12). From these 8 proteins, the structure of 6 proteins was already known with the other two were solved using this complex (Engilberge et al., 2017). The crystallization assays on these proteins were all performed at the HTXLab platform, with 576 screening conditions evaluated. To limit bias in result evaluation, each protein was crystallized in the presence and in the absence of the molecule.

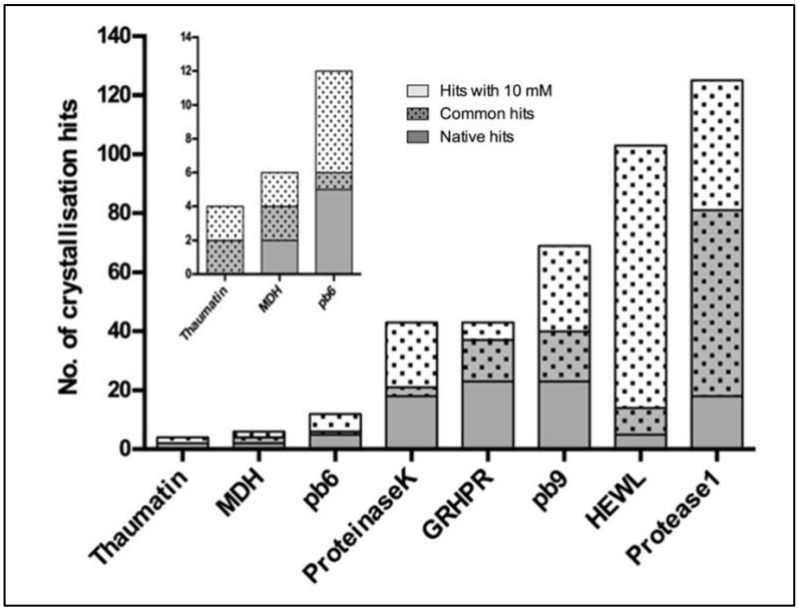


Figure 1-12 *Automated crystallization screening (576 conditions) performed on 8 proteins (drops observed after 34 days). grey: number of hits for native protein, dots: number of hits for protein supplemented with 10 mM of TbXo4 (Engilberge et al., 2017).*

The results of the automated crystallization in the figure above have shown that TbXo4 provides new unique crystallization conditions in all the eight protein models, with the recovery of some of the native protein crystallization conditions.

Also, TbXo4 has proven that its effects on crystallization results in the enlargement of the crystallization diagram of a given protein as illustrated in Figure 1-13 a. This observation has been obtained and validated with several proteins.

Moreover, beyond its nucleating property, the crystallophore has permitted to improve the crystal quality as illustrated in Figure 1-13 b, where two model proteins were crystallized strictly in similar crystallization conditions each. The presence of TbXo4 improve the crystal quality by making them bigger and with better crystal habit.

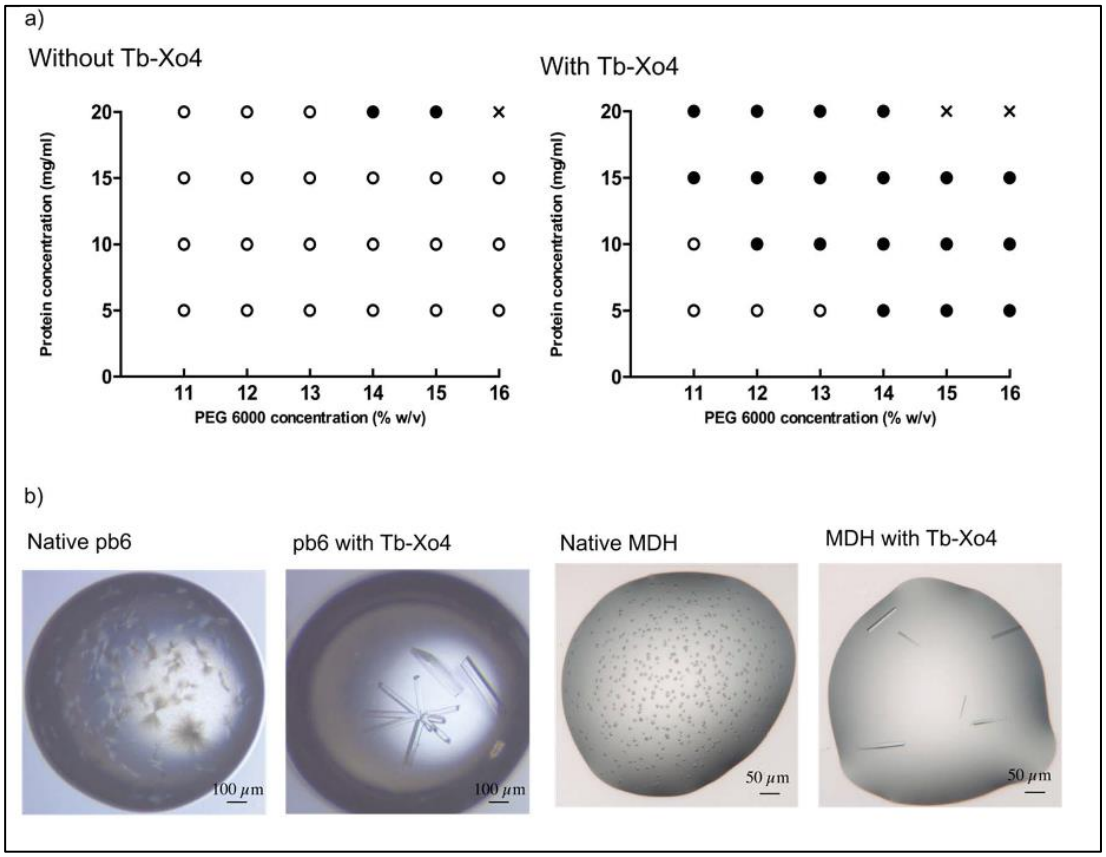


Figure 1-13 *Crystallization diagrams determined for T5-phage pb6 protein in the absence (left) and in the presence (right) of TbXo4 (10 mM). Black dots represent conditions leading to crystals (hits). White dots represent clear drops. Crosses represent drops with protein precipitates. (b) Comparison of the crystals' quality obtained under strictly similar crystallization conditions for the two proteins of unknown structures pb6 and MDH, in the absence and presence of 10 mM of TbXo4 as indicated (Engilberge et al., 2017).*

Furthermore, TbXo4 had proven its capability to crystallize fraction of natively purified proteins from the organism of origin (Figure 1-14), leading to exploitable crystals in size and diffraction quality (Figure 1-14 c-f) (Engilberge et al., 2019).

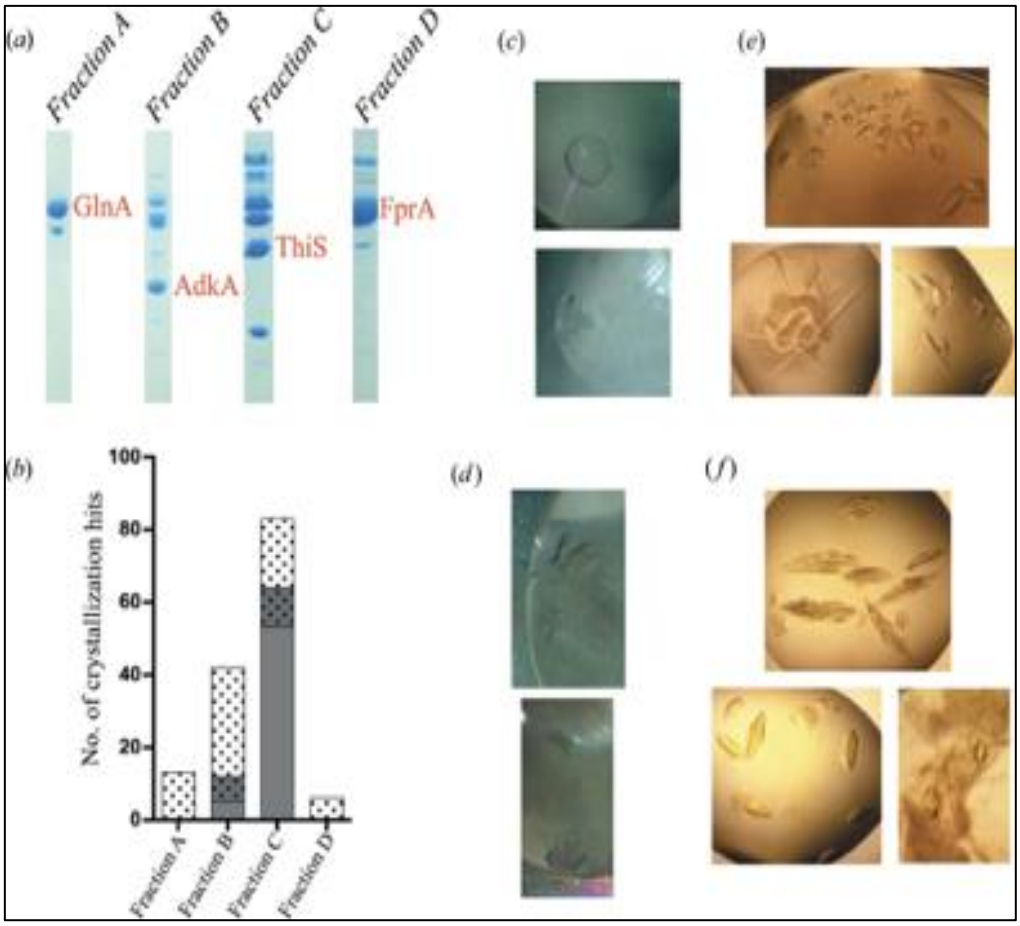


Figure 1-14 *Samples obtained from M. thermolithotrophicus by native purification. (a) SDS-PAGE gels of the purified proteins fractions. (b) 96 Crystallization screening conditions performed on the proteins fractions. The number of unique crystallization hits is depicted in grey for the native protein without TbXo4 and with dots for the protein supplemented with 10 mM TbXo4. Noted that Fraction A and D were only evaluated with TbXo4. (c-f) Crystallization results on fractions A to D in presence of 10 mM TbXo4 (Engilberge et al., 2019).*

In addition, Sylvain had also demonstrated the efficiency of the TbXo4 approach for structure determination, in particular (i) to overcome often-encountered problems such as twinning or low-resolution diffracting crystals, (ii) to promote different crystal packing of the same protein and finally, (iii) to show the compatibility of TbXo4 with serial crystallography approaches (Engilberge et al., 2019).

1.5.2.3 TbXo4 binding mode

To explain and hopefully improve crystallophore properties, the interaction of TbXo4 has been described in different crystallographic structures (Engilberge et al., 2018).

The analyses of the TbXo4 binding mode have been associated to calculations based on the density functional theory (DFT) which allows to figure out the energies governing the interactions between the complex and the protein surface. This preliminary analysis has shown the versatile binding modes of TbXo4, resulting from different balances between electrostatic interactions, hydrogen bonds and hydrophobic interactions.

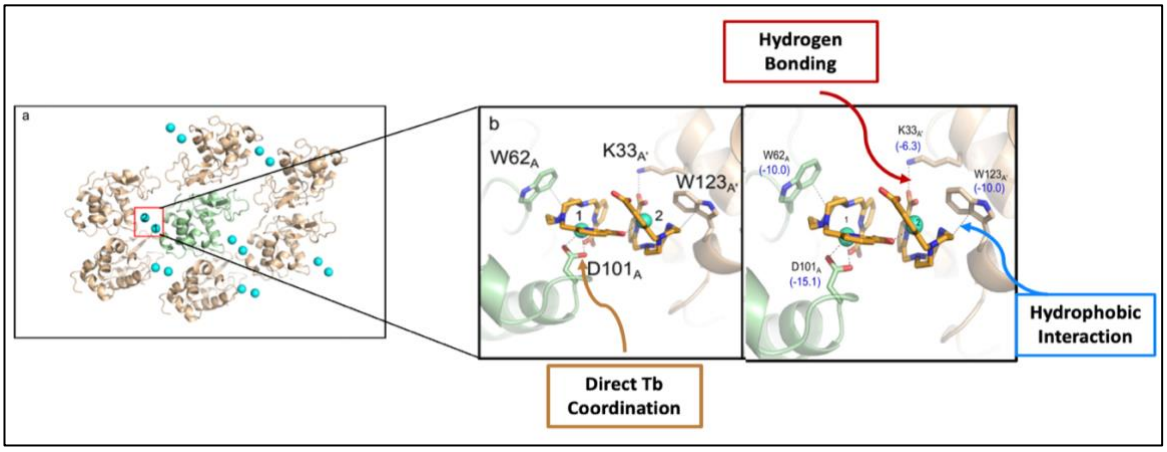


Figure 1-15 *Illustration of TbXo4 dimer interactions in HEWL structure. HEWL was co-crystallized with 100mM TbXo4 a) crystal packing of HEWL in presence of TbXo4 dimer. b) TbXo4 dimer and its interactions in the structure of HEWL. Asymmetric unit content are shown in green and symmetry-related molecules represented in light brown. Values in blue refer to interaction energies evaluated by DFT (in kcal/mol) PDB: 6F2I. (adapted from (Engilberge et al., 2018)).*

An example of TbXo4 interactions are presented in Figure 1-15 in the case of HEWL. This structure of HEWL was obtained with crystals co-crystallized with 100 mM TbXo4. Two TbXo4 were found associated as a dimer in the asymmetric unit, specifically at the interface of three molecules of HEWL (Figure 1-15 a). This association was made by the interaction of a carbonyl unit of a picolinate arm from the first crystallophore with the metal of the second lanthanide complex (Figure 1-15 b). Also, the first complex coordinates to the carboxy oxygens of an aspartate at the position 101 (D101) of the first molecule (labelled A), and the second complex forms a hydrogen bond between the carbonyl group of the picolinate ligand and the ammonium moiety of a lysine (K33) from a symmetry-related molecule (label A') (Figure 1-15 b-c). In complement, the two TbXo4 molecules forms hydrophobic interactions with two tryptophan residues $W62_A$ and $W123_{A'}$ (Engilberge et al., 2018).

In addition to this structure, 5 other crystal structures, detailed in (Engilberge et al., 2018), have demonstrated the versatile interaction between TbXo4 and the protein, as it adapts to the chemical environment such as the crystallization components and the size and chemical nature of the binding pocket, in order to optimize its contacts with the protein of interest through several supramolecular interactions.
Therefore, TbXo4 has shown that it is an all-in-one, easy-to-use as its overcomes the two main hurdles of macromolecular crystallography: obtaining high quality single crystals and solving the phase problem.

1.6 Aim of the thesis

Despite all the advantageous properties described previously, the different studies undergone on TbXo4 had shown that it is not 100% successful in term of crystallization, leaving space for improvements.
Therefore, the next step is to decipher the origin of the nucleating properties of TbXo4 and its mechanism of interaction on the surface of proteins, in order to increase TbXo4 success rate.

For that, our approach has been two-fold:

- First approach is from the chemical side. It consists in exploiting the chemistry offered by lanthanide complexes, in particular TbXo4. In this respect, several TbXo4 variants were synthetized by addition of moieties either on the macrocycle nor on the picolinate arm. In order to evaluate the influence of these addition on the crystallization, the nucleating properties of TbXo4 variants were determined using the HTX Lab crystallization platform (https://embl.fr/htxlab/) on a panel of 3 commercial proteins. This approach has shed new light on the complex issue of protein crystallisation but has also allowed to optimize the use of TbXo4.

- Second approach is a biological approach that consists of analysing the interaction sites between protein surfaces and TbXo4, by crystal structure analysis and theoretical calculations. This approach is complemented with a mutagenesis approach (point mutations) to target the protein motifs potentially involved in the binding of TbXo4. This second approach has allowed to

determine the predominant role of certain amino acids in the crystallization induced by TbXo4.

These two approaches will be presented in chapters 3 and 4, respectively. The manuscript will also contained a chapter (chapter 2) dedicated to the experimental protocols used to prepare the samples and to perform the experiments and will be ended with a general conclusion and potential perspectives of the present work.

CHAPTER 2

MATERIALS AND METHODS

2. MATERIALS AND METHODS

In this chapter, I will present the details about the proteins used in this thesis; their preparation, expression and purification as well as the crystallization screening kits used and the protein structures determination protocol.

2.1 Preparation of commercial protein's solutions

To study TbXo4 variants, Hen Egg-White Lysozyme from *Gallus gallus* (Roche reference: 10837059001), Proteinase K from *Parengyodontium album* expressed in *Pichia pastoris* (Roche reference: 03115879001) and Thaumatin from *Thaumatococcus daniellii* (Sigma reference: T7638) (Table 2-1) were prepared by adding milli-Q water to the weighed lyophilized protein's powder. Protein concentrations were verified by spectrophotometric measurement at 280 nm using a Thermo Scientific Nano Drop 2000 spectrophotometer.

2.2 Recombinant Proteins

2.2.1 Expression

For Adenylate kinase from *Methanothermococcus thermolithotrophicus (Mth*AdkA*)*, the proteins of interest (Table 2-1) were expressed in competent *Escherichia coli* (*E. coli*) BL21DE3 cells, after being transformed with a Pet41a plasmid containing the gene encoding for the protein of interest which was provided by the company GeneCust.

Transformed cells were placed on a Petri dish in which culture medium of LB broth with agar (Lennox) from Sigma at 35 g/L, enriched with 50 µM of the specific antibiotic of the inserted vector, was poured and left overnight at 37 °C in a dry incubator.

For the first expression, the pre-culture was prepared by plating a colony from the petri dish culture into 50 ml of autoclaved LB medium from Sigma at 20 g/L and enriched with final concentration of 50 µM of specific antibiotic of the inserted vector.

After an overnight incubation of the pre-culture at 180 rpm at 37 °C, a cell stock stored at -80 °C, with 700 µl of pre-culture homogenised with 30% of final concentration of sterile glycerol was prepared. Then, 2 litre of LB medium 20 g/L, enriched with 50 µM of the specific antibiotic, was inoculated with 50 ml of pre-culture. When the optical density (OD) of the culture reaches 0.6-0.8, protein expression was induced with a final concentration of 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG). The culture was again incubated for 4 hrs at 37 °C under agitation at 180 rpm. The culture volume is then centrifuged at 4000 rpm for 20 min at 4 °C using a Fiberlite F9-6x1000 LEX (PN:096-061075) rotor in a Thermo Scientific Sorvall LYNX 6000 centrifuge. The collected bacterial pellets were resuspended with 50 ml of lysis buffer (20 mM Tris-HCl, 50 mM NaCl, 0.1% Triton X-100, pH 7.5) and stored in 50 ml tubes at -80 °C.

Table 2-1: *List of proteins, Vectors and Antibiotic Resistance (AR), Species of origin, expression strain, correlation coefficient(ε) and their theorical molecular weight.*

Proteins	Vector, AR	Species of origin	Expression strain	Extinction Coefficient (ε)	Molecular Weight (Da)	pI	%Asp & Glu	%Arg,His& Lys
Thaumatin I	Native	Thaumatococcus daniellii	X	29420	22205.03	8.32	9.17	11.11
Proteinase K	Native	Parengyodontium album	X	36580	28906.84	8.25	6.45	7.16
HEWL	Native	Gallus gallus	X	37970	14313.14	9.32	6.97	13.17
TEWL	Native	Meleagris gallopavo	X	39460	14209.07	9.41	6.2	13.17
QEWL	Native	Coturnix japonica	X	39460	14366.24	9.28	6.97	13.17
PEWL	Native	Pavo cristatus (Indian peafowl)	X	39460	14422.31	9.43	6.97	13.95
MthAdkA-wt	pET41a, KanR	Methanothermococcus thermolithotrophicus	*E. coli*	11460	21460.78	5.37	14.58	12.5
MthAdkA-M1	pET41a, KanR	Methanothermococcus thermolithotrophicus	*E. coli*	11460	21402.75	5.55	14.06	12.5
MthAdkA-M2	pET41a, KanR	Methanothermococcus thermolithotrophicus	*E. coli*	11460	21416.77	5.56	14.06	12.5
MthAdkA-M3	pET41a, KanR	Methanothermococcus thermolithotrophicus	*E. coli*	11460	21459.80	5.55	14.06	12.5
MthAdkA-M4	pET41a, KanR	Methanothermococcus thermolithotrophicus	*E. coli*	11460	21446.76	5.36	14.58	12.5
MthAdkA-M5	pET41a, KanR	Methanothermococcus thermolithotrophicus	*E. coli*	11460	21459.80	5.55	14.06	12.5
MthAdkA-M6	pET41a, KanR	Methanothermococcus thermolithotrophicus	*E. coli*	11460	21454.88	8.78	11.45	12.5

2.2.2 Purification protocol of recombinant proteins (MthAdkA)

After thawing the recovered bacterial pellets, 0.25 mg/ml lysozyme (Sigma), 0.05 mg/ml DNase (Roche), 0.2 mg/ml RNAse (Roche), 1 tablet of protease cocktail "Complete" (Roche) and a final concentration of 20mM $MgSO_4$ were added. Then, the cell lysates were ultrasonicated 5 times for 45 s with 30 s breaks in ice. As MthAdkA is derived from a thermophilic organism, a heat shock at 65°C for 20 min of the cell extract was performed (SN_b on Figure 2-1). The heat shock step is considered as the first purification step that removes the maximum of contaminating proteins from *E. coli*. The denatured proteins from the heat shock were then separated from the soluble proteins by centrifugation at 18 000 rpm for 45 min at 4 °C using JA-20 rotor in a Beckman avanti^{TM}J-20I centrifuge.

Then, the supernatant (SN_a) is collected and filtered with a 0.22 µm filter.

After that, a strong anion exchange column (6ml Ressource Q, GE Healthcare) was used as an intermediate step. The column was first equilibrated with 30ml of Buffer A (20 mM Tris–HCl (pH 7.8), 50 mM NaCl, 50 mM KCl, 10 mM $MgCl_2$), then the filtered supernatant was injected to the column and washed with 10 column volume (CV) of Buffer A to remove contaminating proteins with little interaction with the resin. Absorbed proteins (PF) were eluted by using a NaCl step-gradient of 7 CV, from 50mM to 500 mM of NaCl using Buffer B (20 mM Tris–HCl (pH 7.8), 1 M NaCl, 50 mM KCl, 10 mM MgCl2) (Figure 2-1). The final purification step was performed on a size exclusion column (Superose 12, GE Healthcare) previously equilibrated with Buffer A (Figure 2-2).

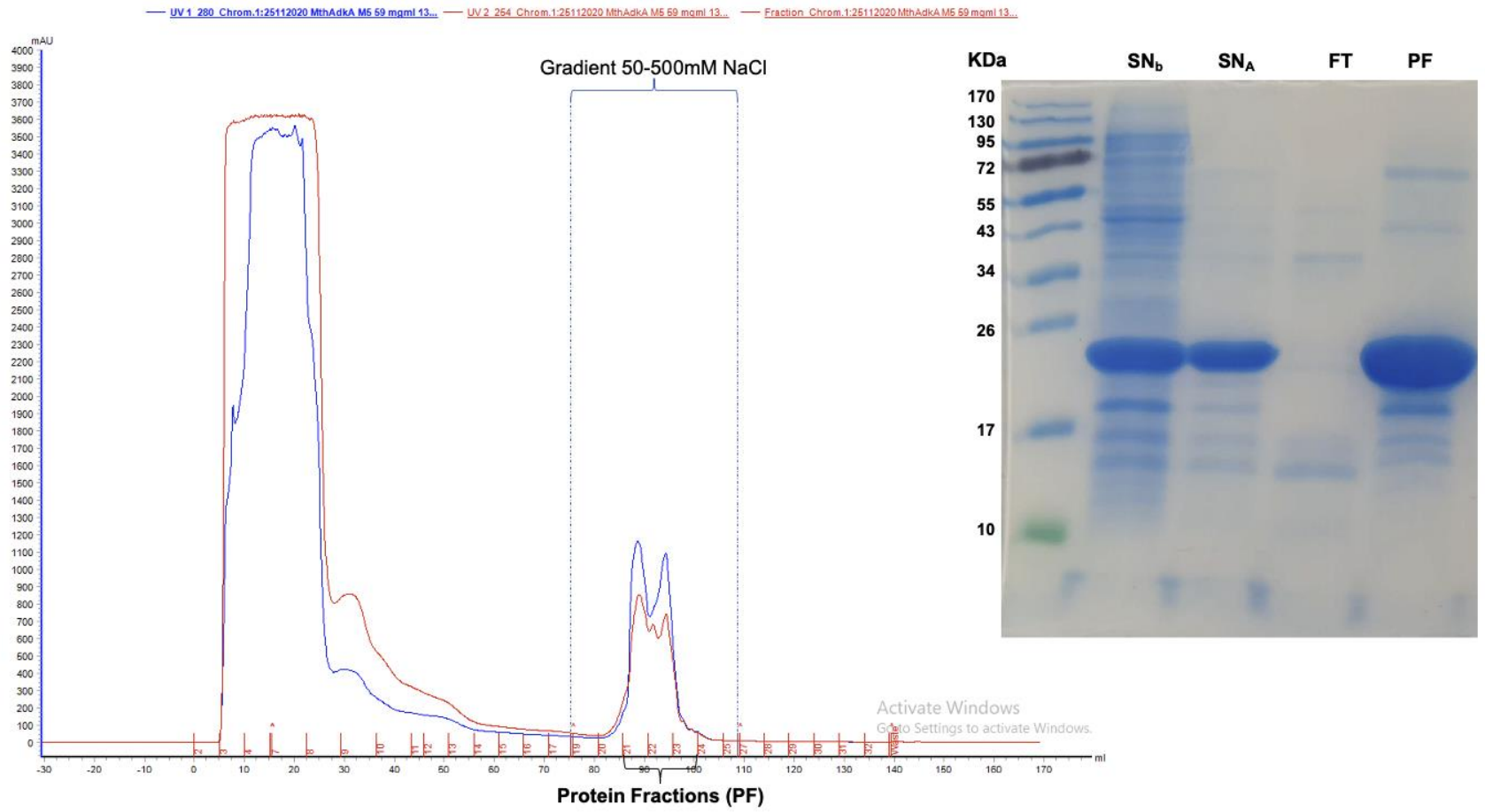


Figure 2-1 *Example of elution chromatogram of MthAdkA after an ion exchange column and its SDS-PAGE gel on the right. The blue and red curves represent the U.V. absorbance at 280 nm and 254 nm, respectively. SN_b: Supernatant before heat, SN_a: Supernatant after heat, FT: Flowthrough and PF: Protein Fractions.*

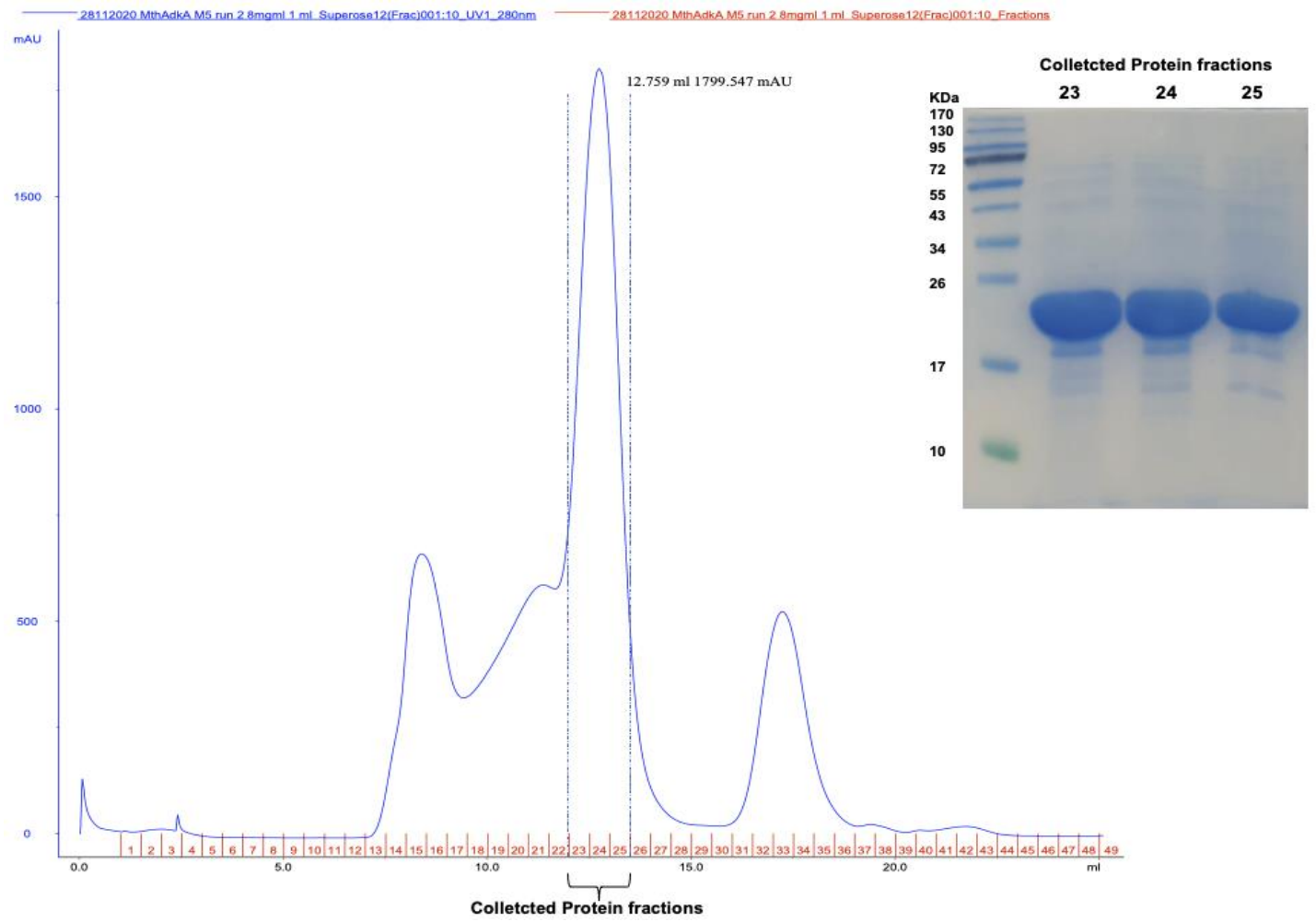


Figure 2-2 *Example of elution chromatogram of MthAdkA after gel filtration and its SDS-PAGE gel on the right. The blue and red curves represent the U.V. absorbance at 280 nm and 254 nm, respectively. Collected Protein fractions i.e. protein fractions used for crystallization.*

Next, a 15% SDS-PAGE gel (Figure 2-2) with Coomassie Blue revelation was run to verify the fraction containing the protein and its purity, the latter was also confirmed by mass spectrometry on the IBS mass spectrometry platform.

Finally, pooled fractions of the protein were concentrated using centrifugal filter Amicon® Ultra-15 10K with regenerated cellulose membrane. Before crystallization, the concentration was verified by spectrophotometric measurement at 280 nm, using a Thermo Scientific Nano Drop 2000 spectrophotometer, with the molar extinction coefficient (ε) of the protein determined from the protein sequence by the PROTPARAM software (http://web.expasy.org/protparam).

2.3 Purification protocol of Egg-White Lysozyme

Fresh eggs were purchased from local stores for Quail eggs, local farms for Turkey eggs, national zoos or reserves for Peacock eggs.

Jacques Coves, CNRS Research director in our Team ELMA, carried out all the purifications of Egg-White Lysozymes. The following protocol is the one he set up.

The shells were manually broken or cut out then the Egg-White was separated from the yolk. Then, 100 ml of the Egg-White was mixed with 250-300 ml of 0.1 M Glycine at pH 9.3 (Buffer A') and stirred for 1 hour. After a centrifugation of the mixture at 18 000 rpm for 30 minutes at 4 °C using a JA20 rotor, the supernatant was loaded onto a 60 ml CM Sepharose Fast-Flow column at about 2 $ml.min^{-1}$, after the column was equilibrated with 3 volumes of Buffer A'. The flow-through was discarded and the column was washed with Buffer A' until the absorbance at 280 nm was close to 0. Lysozyme and the remaining contaminant proteins were eluted in one step by two volumes of Buffer A' supplemented with 1 M NaCl.

Ammonium sulfate was added at a final saturation of 80% and the mixture was stirred gently overnight at 4°C. Precipitated proteins were recovered by centrifugation. The pellets was dissolved in a minimal volume of 50 mM Tris-HCl pH=7.5 50 mM NaCl (Buffer B'), clarified by centrifugation and 2 ml of supernatant were loaded at 0.8 $ml.min^{-1}$ onto a Superdex-75 16/60 prep grade (120 ml). The column, previously equilibrated with Buffer B' at 0.8 $ml.min^{-1}$, was eluted by the same buffer at the same flow rate (Figure 2-3). Next, SDS-PAGE gel with Coomassie Blue revelation was run to verify the fraction containing the protein and its purity, the latter was confirmed by

mass spectrometry on the IBS mass spectrometry platform. Before crystallization, fractions containing Lysozyme were pooled and concentrated, using a centrifugal filter VIVASPIN 20 - 5K PES from Sartorius, until a concentration of 20 $mg.ml^{-1}$ is reached.

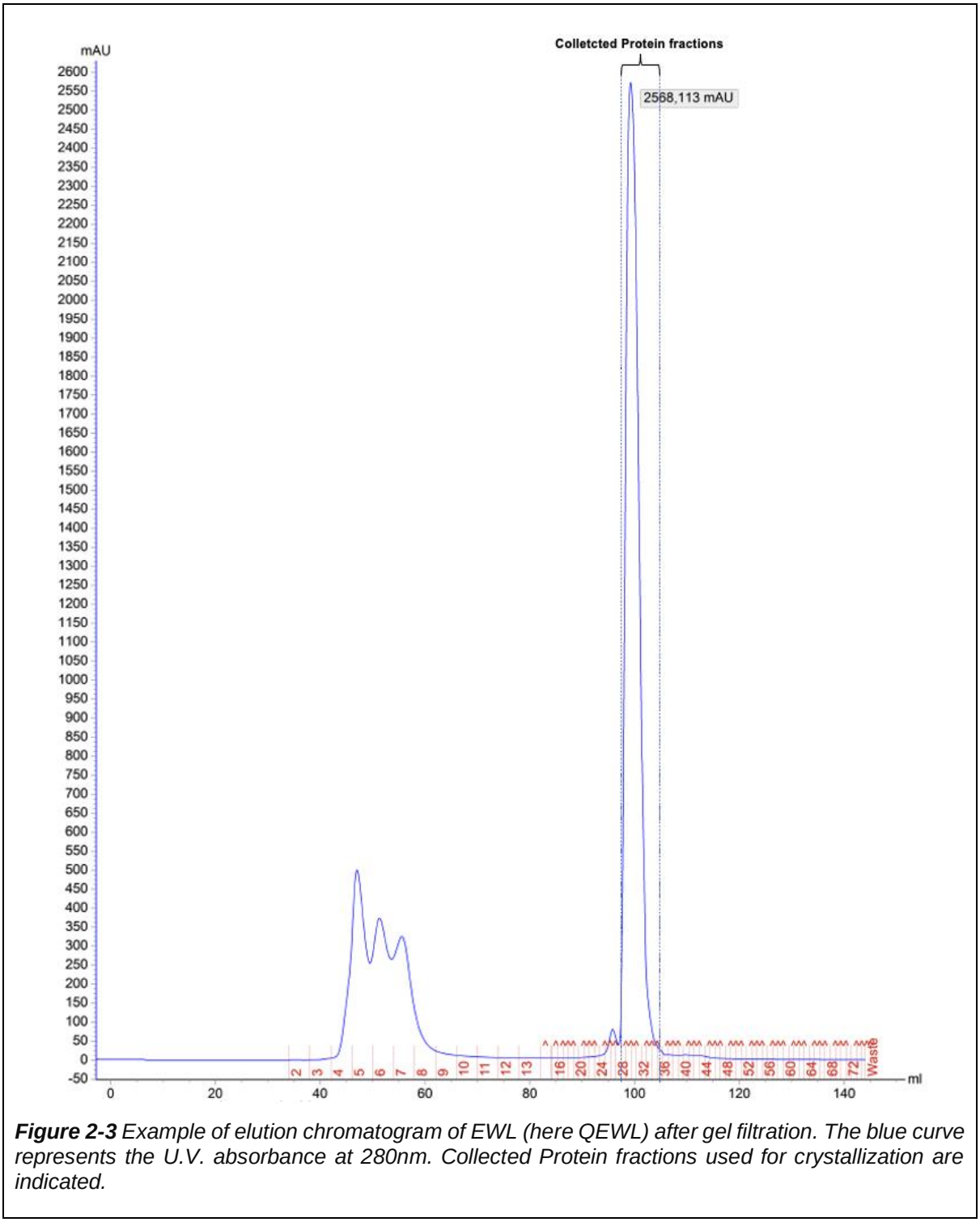


Figure 2-3 *Example of elution chromatogram of EWL (here QEWL) after gel filtration. The blue curve represents the U.V. absorbance at 280nm. Collected Protein fractions used for crystallization are indicated.*

2.4 Crystallization screening

In Grenoble, we have access to the automated screening platform HTX lab at EMBL with a robotic systems nanovolume dispenser (Mosquito TTP Labtech). During this thesis, the automated screening experiments were performed, for each protein, using the sitting drop technique with a standard drops set up, where a volume of 100 nl of protein or protein/Xo4 is mixed with 100 nl of precipitant solution and this drop is then equilibrated against a well of 45 µl. The evaluations, for each protein, were performed on 6 screening kits for a total of 576 different crystallisation conditions (Table 2-2), which requires a total of only 75 µl of protein or protein/Xo4 solution.
To improve the crystals, I based my work on their shapes. I favoured single crystals with well-defined edges.

Table 2-2: *Screening kits used during this thesis.*

Screen Supplier	**Screen name**
Qiagen/Nextal	Classics Suite (Qiagen)
Molecular Dimensions	JCSG +
Molecular Dimensions	PACT
Qiagen/Nextal	PEGs I
Hampton Research	Salt Grid
Rigaku Reagents	Wizard I & II

2.5 Protein structure determination by X-ray crystallography

2.5.1 Diffraction data recording

During this thesis, diffraction data were recorded at SOLEIL synchrotron or ESRF Synchrotron after the crystals were flash-cooled in liquid nitrogen at 100K after treatment, for a few seconds, in a cryoprotective solution. Diffraction data were recorded at L_{III} absorption edge of terbium (1.648 Å) to exploit at best the anomalous contribution of the lanthanide atom.

2.5.2 Diffraction data processing

A dataset consists of a set of X-ray diffraction patterns which is a distorted image of the reciprocal lattice.

To determine the corresponding intensities I_{hkl} of each of the reflections in these diffractions patterns, we use the XDS software (Kabsch, 2010) or the XDSME package (P. Legrand, 2017, https://github.com/legrandp/xdsme) (Legrandp et al., 2019). The whole integration process involves several steps. The first step is the indexing which is realized by the IDXREF subroutine. Like its name, this indexing is mainly done on the most intense diffraction spots. It permits to assign a Miller index (hkl) for each observed reflection from the diffraction pattern, determining the orientation of the crystal and the crystal lattice parameters. Also it gives a first indication of the symmetry of the crystal (Bravais lattice).

Second step is the integration of the diffraction data performed by the INTEGRATE subroutine. The integration consists of determining the positions of the diffraction spots related to the diffraction pattern, based on the parameters determined in the previous step, and then evaluating and integrating the intensities measured on these spots, therefore estimating the value of the corresponding intensity I_{hkl} and the associated standard deviation σ_{hkl} .

Last step in XDS is the scaling which is performed by the CORRECT subroutine. This step allows to obtain a list of reflections with their average and corrected intensity by taking into account the equivalent reflections, from symmetries associated reflections with the space group or from reflections measured several times. CORRECT also provides a first proposal for a possible space group which can be verified later by the program POINTLESS(Evans, 2005).

After this step from XDS, a set of statistics is obtained which allows to evaluate the data quality and the choice of maximum resolution depending on the following indicators (Karplus & Diederichs, 2012):

1. Signal-to-noise ratio ($I/\sigma(I)$) with a cutoff of 2.

2. R_{sym} that indicates the agreement between the symmetrically equivalent intensities and their average.
3. The correlation coefficient CC1/2 between two dataset split in in random ways. The CC1/2 reflects the correlation between the two. This parameter estimates the proportion of signal over noise that contributes to the variation in the data with a cutoff of 0.5.
4. Completeness of the data > 90%.

XDS steps can be repeated several times, with possibilities of images exclusion that worsen the statistics without improving completeness, until the refinement of the crystalline and experimental parameters.

To calculate the amplitudes of the structure factors from the intensities, a script written in CSHELL was used. This script convert the data from XDS to the appropriate format mtz using COMBAT and SORTMTZ programs, then the SCALA program from CCP4 (Evans, 2005) for scaling and merge the integrated intensities and finally the TRUNCATE program to produce the structure factors amplitudes. Other than the indicators for data quality mentioned above, another indicator R_{pim}, which takes into account the redundancy of the data, is present in the log file generated by SCALA and use to evaluate data quality. Also, the mtz file produced by the program TRUNCATE is used to check the quality of the data as it shows if there is presence twinning and pseudo-symmetry translation in the data.

After this step, an estimation of the number of molecules presented in the asymmetric unit was done by estimating the Matthews coefficient using a program with the same name available in the CCP4 (Kantardjieff & Rupp, 2003).

Structures in this thesis were determined by molecular replacement using models with > 90% sequence identities. PHASER software from CCP4 (McCoy et al., 2007) was used for the determination of the structures.

The next step is the refinement of the models which was done with the PHENIX package (phenix.refine) (Adams et al., 2010) associated with manual reconstruction of the models with COOT (Emsley & Cowtan, 2004). A typical structure refinement consists in several rounds described hereafter. The first round of model refinement

consisted in a rigid body refinement followed by a simulated annealing and the refinement of atomic positions and thermal agitation factors (B-factors). After this step, the model was reconstructed and checked manually in COOT. After this first round of refinement, the terbium atoms (Tb) were placed thanks to an anomalous Fourier synthesis calculated with the model obtained after the first round as phase source. The occupancy rate of Tb was adjusted so that the thermal agitation factor of the Tb atom is close to that of the coordinating atoms of the amino acids with which it interacts. Second round of refinement consisted in the refinement of atomic positions and B-factors as well as an automatic search for water molecules. The addition of water molecules tends to strongly reduce the R_{factor} and R_{free}, statistic indicators for the quality of the refinement. At this stage, Xo4 ligand or its variants were placed into the density using COOT and with the same occupancy rate as their Tb atom. Xo4 and variant ligand coordinates and refinement restraints were generated using eLBOW in Phenix, by providing their SMILES string. Also, ions (Na^+ and Cl^-) and molecules present in the composition of the mother liquor and the cryo-freezing solution were positioned. Third and fourth round of refinement consisted in the refinement of the positions of the atoms and the agitation factors always with the automatic search for water molecules. During the fourth round, I also added TLS ("Translation, Liberation Screw") refinement, even if it was not performed in all the structures determined in this thesis. TLS groups were defined using the phenix.find_tls.groups program.

Fifth round of refinement was the optimization of the weights of the atomic positions and agitation factors, search for water molecules and addition of TLS group.

After these different rounds of refinement, the structures were analysed with MOLPROBITY (Chen et al., 2010) and the conformation of amino acids with unfavourable geometry and/or contacts was manually improved in COOT, and water molecule positions and occupations were adjusted. Then a final run of refinement was done with automatic search of water molecule off.

CHAPTER 3

EVALUATION OF TBXO4 VARIANTS

3. EVALUATION OF TBXO4 VARIANTS

As previously described in the introduction, the various studies undergone on TbXo4 show that it is not 100% successful, in term of crystallization. Hence, it remains necessary to understand TbXo4 nucleation mechanism and its behaviour at the surface of proteins in order to increase its success rate.

One approach explored during this thesis is to exploit the chemistry offered by lanthanides complexes, in particular the Xo4 scaffold, to decipher the potential influence of chemical moieties on the nucleating properties of the crystallophore. For that a series of molecules based on Xo4 and named variants have been synthetized by our collaborators at ENS-Lyon.

In the present chapter, we will first address an issue that we had encountered, illustrating some difficulties associated with methodological developments, but which finally turned out to be informative on the TbXo4 nucleating effect. Next, we will present the results of the nucleating properties of several variants on three commercial proteins and their comparison with TbXo4. Then, we will describe the binding mode of the most efficient variants and compare it to the one observed for TbXo4, by exploiting diffraction data of Hen Egg White Lysozyme (HEWL), as model protein, in presence of these compounds. Finally, in the context of our collaboration with the ENS-Lyon team, molecular dynamics simulations of HEWL in presence of TbXo4 and the most efficient variants were carried out by Elise DUMONT and Natacha GILLET.

3.1 TbXo4 reproducibility issues

The beginning of this thesis was characterized by some concerns. First, we tried to reproduce the sample preparation protocol used by S. Engilberge (Engilberge et al., 2017) as well as some crystallizations that had been conducted. However, we first observed the precipitation of some proteins during mixing with TbXo4 powder, precipitation that had not been described for these proteins, previously (Figure 3-1 a & b). Luckily, this precipitation turned out, in some cases, to be crystalline precipitates (Figure 3-1 a). It should be mentioned that we used new batches of TbXo4 molecule, resulting from an improved synthesis/purification protocol compared to the batch used by S. Engilberge (noted TbXo4-SE hereafter). As a consequence, the nature of the

counter-ion had changed from a carbonate (HCO_3^-) to a chloride in this new protocol. As observed in Figure 3-1 b, the addition of HCO_3^- to the protein-TbXo4 precipitate dissolve the mixture, turning it into a clear drop (Figure 3-1 b right) . Therefore, HCO_3^- is considered to have an effect on the nucleation properties of TbXo4, which will be discussed at the end of this section.

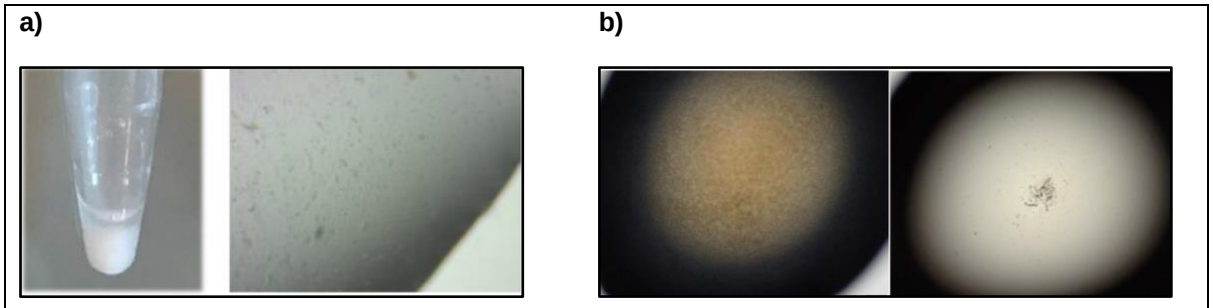


Figure 3-1 *TbXo4 reproducibility test a) Crystalline precipitates of pb9 protein from phage T5 after mixing with 10 mM TbXo4 in the Eppendorf tube (left) and the microscopic view of the mixture (right) b) precipitate of catalase from bovine liver after mixing with TbXo4 to a final concentration of 1 0mM (left), clear drop after adding HCO_3^-(right).*

Moreover, in order to characterize the behaviour of these new batches of TbXo4, we set the crystallization assays of 3 commercial proteins, Hen Egg-White Lysozyme (HEWL), ProteinaseK and Thaumatin, by mixing the protein solutions (at 20 mg/ml) with TbXo4 powder to a final concentration of 10 mM. The crystallization assays were performed similarly to the ones by S. Engilberge (i.e. 1+1 protocol: 1 volume of protein solution + 1 volume of well solution) (Engilberge et al., 2017) using an automated crystallization screening platform (HTX-Lab, EMBL, Grenoble). The crystallization assays results using the new TbXo4 batches as well as the results obtained by S. Engilberge (Engilberge et al., 2017), which were obtained using the batch of TbXo4 from the previous synthesis protocol (TbXo4-SE), are presented in Figure 3-2. By comparing the number of crystallization hits (i.e. the number of crystallization conditions with well-defined crystals) of the 3 commercials proteins, in presence of TbXo4 issued from the two synthesis protocols, we observe an increase of crystallization hits with the new batches of TbXo4 compared to the number obtained with TbXo4-SE. This clearly shows a higher nucleating effect of TbXo4 issued from the new synthesis protocol.

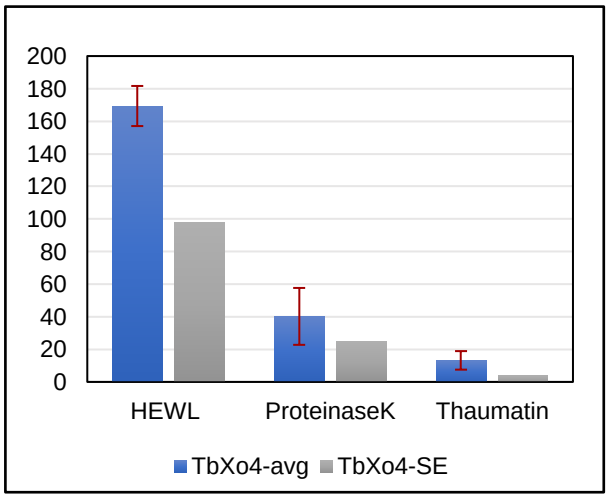


Figure 3-2 *Number of crystallization hits of three commercial proteins, HEWL (Hen Egg-White Lysozyme), ProteinaseK and Thaumatin in presence of TbXo4 batches, using 1+1 crystallization protocol test. In (blue) the average (avg) number of crystallization hits using the new TbXo4 batch and its standard deviation in red error bars and in (grey) the number of crystallization hits from TbXo4-SE batch (Engilberge et al., 2017).*

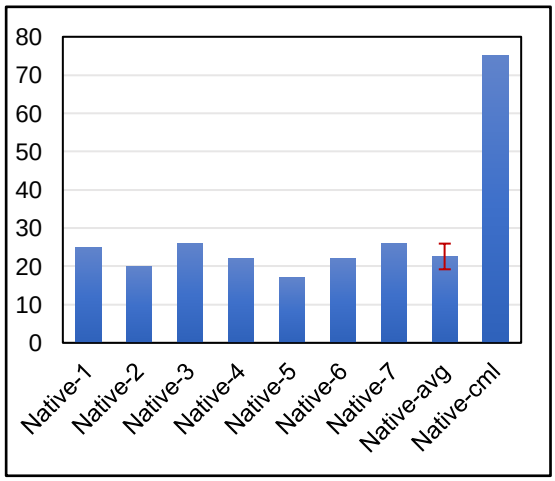
Noted that the results presented in Figure 3-2 in blue are the average number of crystallization hits obtained from several crystallization assays (i.e. the sum of the number of crystallization hits from the same type of crystallization trials of a protein divided by the number of crystallization trials) of the commercial proteins in presence of 10 mM TbXo4 (Figures 3-3, 3-4, 3-5 b). Indeed, through this thesis, I have performed several crystallisation trials on these three commercial proteins, without and with 10 mM TbXo4, to check the crystallization reproducibility as we used different protein lots and/or different TbXo4 batches, all issued from the new purification protocol. The results of these several crystallization trials will be described below.

For HEWL, the crystallization trials for the protein without and with TbXo4 are shown in Figure 3-3 a & b, respectively. The table in Figure 3-3 c shows the protein lot and TbXo4 batch used in each crystallization assay. Two protein lots from two different suppliers were used for the crystallization assays: protein from Sigma-Aldrich was used for Native-1 and TbXo4-1,2,3 assays and from Roche for the rest. Also, two TbXo4 batches were exploited for HEWL crystallization assays with TbXo4 batch 1 for TbXo4-1,2 assays and batch 2 for the rest (Figure 3-3 c). Noted that the two TbXo4 batches (1 & 2) were obtained from the same purification protocol, as already mentioned.

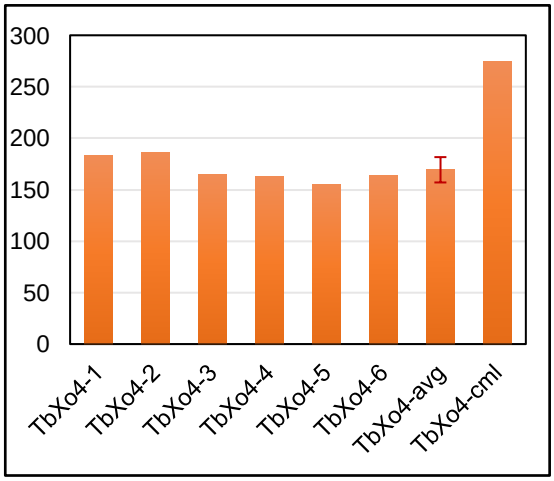
The results from the native crystallization assays of HEWL show an apparent reproducibility of the crystallization assays with a variability between 10-25% versus the average value. This variability can be explained by the stochastic nature of the crystallization process, the protein lots as well as the storage conditions of the commercial kits before the experiments (time and temperature), which may have an effect on the pH of the crystallization screens mixture (Wilson et al., 2020).

The results of the HEWL-TbXo4 crystallization assays show, as expected from the work of S. Engilberge, an increase in HEWL nucleation compared to the native crystallization assays. Also, reproducibility of the crystallization experiments with a minimal variability of ~10% with respect to the mean value is observed with TbXo4, which is lower than the value observed with the native assays. Thus, TbXo4 decreases the variability of crystallization assays with HEWL.

a)



b)



c)

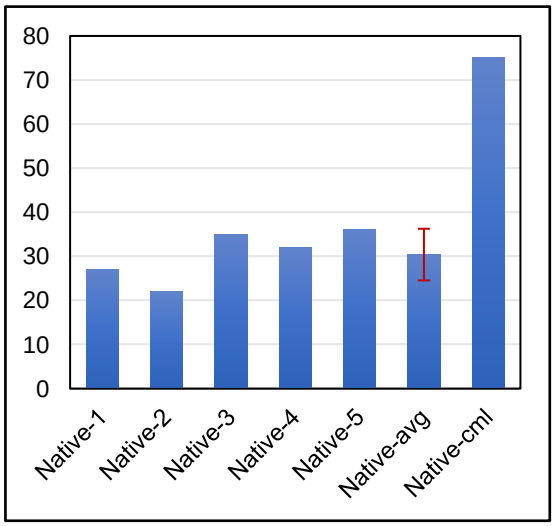
	Crystallization assays		
Protein lots	Native	TbXo4	
		Batch 1	Batch 2
Sigma-Aldrich	1	1 & 2	3
Roche	2, 3, 4, 5, 6 & 7	-	4, 5 & 6

Figure 3-3 *Number of crystallization hits of HEWL (a) in absence blue and (b) in presence (orange) of TbXo4, using 1+1 crystallization protocol test, the average(avg) and its standard deviation(error bars in red) and cumulative number (cml) of all the assays. c) Origin of protein lot and TbXo4 batch for each crystallization assay.*

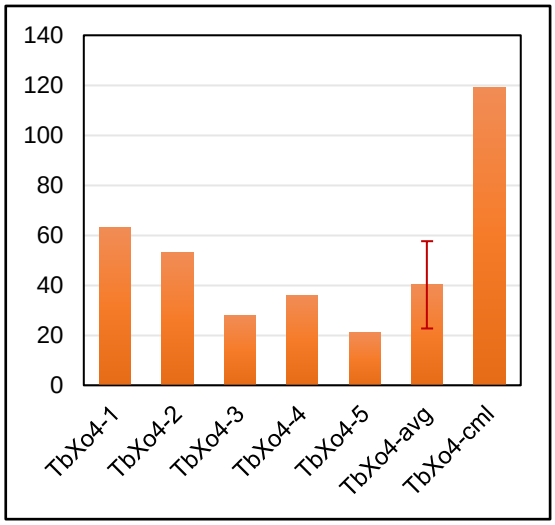
For ProteinaseK, the crystallization trials for the protein without and with TbXo4 are shown in Figure 3-4 a & b, respectively. The table in Figure 3-4 c shows the protein lot and TbXo4 batch used in each crystallization assay. Three protein lots from the same supplier Roche were used, lot 1: TbXo4-1,2; lot 2: TbXo4-3 and lot 3: All the native assays as well as TbXo4-4,5 and two TbXo4 batches were used with batch 1 used in TbXo4-1,2 and batch 2 for TbXo4-3,4,5 (Figure 3-4).

The results of the native ProteinaseK crystallization assays show an apparent reproducibility of the crystallization assays with a variability between 15-30% versus the average value. However, the results of ProteinaseK-TbXo4 crystallization assays, show a reproducibility of the crystallization experiments with a significant variability between 35-50% versus the average value, which is higher than the value obtained with the native assays. Although the average number of crystallisation hits of the ProteinaseK-TbXo4 crystallisation assays is higher than that of the native assays, we cannot specify whether TbXo4 increases the nucleation of ProteinaseK, due to the high variability presented. For the native assays, the dispersion of the number of hits for each assay can therefore be explained mainly by the stochastic nature of protein crystallization, since we used the same batch of protein. As for the crystallisation assays with TbXo4, the dispersion of the number of hits can be explained by quality differences in protein batches as well as the stochastic nature of protein crystallization, again. It should be indicated that the results shown in Figure 3-4 b does not take into account the statistic of an assay on a fourth lot of protein. Indeed, this lot led to a very high number of hits with TbXo4 (173 crystallization hits) compared to the average numbers observed. Thus, we did not include the result of this lot in our study. Which from this comes our conclusion about the major effect of the protein lots on Proteinase K crystallization assays statistics.

a)



b)



c)

		Crystallization assays		
Protein lots		Native	TbXo4	
			Batch 1	Batch 2
Roche	Lot 1		1 & 2	
	Lot 2			3
	Lot 3	1, 2, 3, 4 & 5		4 & 5

Figure 3-4 *Number of crystallization hits of ProteinaseK in absence (blue) and in presence (orange) of TbXo4, using 1+1 crystallization protocol test, the average(avg) and its standard deviation(error bars in red) and cumulative number (cml) of all the assays. c) Origin of protein lot and TbXo4 batch for each crystallization assay.*

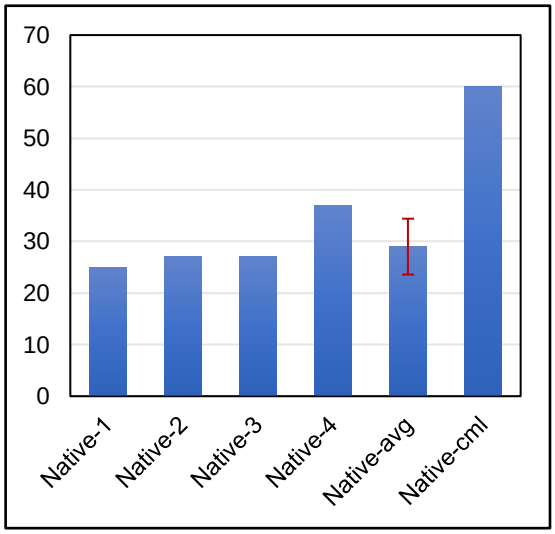
As for Thaumatin, the crystallization assays for the protein without and with TbXo4 are shown in Figure 3-5 a & b, respectively. Similarly to HEWL and ProteinaseK, the table in Figure 3-5 c shows the protein lot and TbXo4 batch used in each crystallization assay. Two protein lots from the same supplier Sigma-Aldrich were used, lot 1: TbXo4-1,2,3 ; lot 2: TbXo4-4; with same TbXo4 batch used for the crystallization assays with TbXo4 (Figure 3-5 c).

In the first instance, by observing the two bar charts of Thaumatin crystallization assays in the absence and presence of TbXo4, we find that TbXo4 decreases the nucleation of Thaumatin instead of increasing it. This shows that the efficiency of TbXo4 is protein dependent.

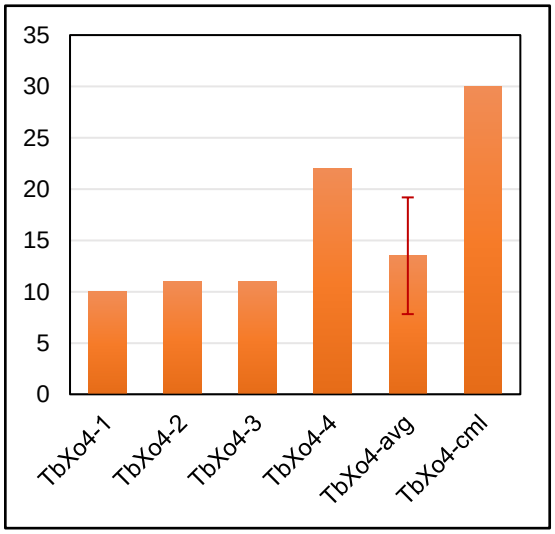
Moreover, the results from native crystallization assays of Thaumatin show an apparent reproducibility of the crystallization assays with a variability between 10-20% versus the average value. However, the results of Thaumatin-TbXo4 crystallization assays, show a reproducibility of the crystallization experiments with a significant variability between 25-40% versus the average value, which is higher than the value obtained with the native assays.

Therefore, the low dispersion of the number of hits with and without TbXo4 for the first three assays can be explained obviously as for HEWL and ProteinaseK, e.g. the nature of the crystallization process, as they were performed with the same lot of protein. However, after changing the protein lot, there was an increase in the number of crystallization hits of Thaumatin-native. Same goes for the fourth crystallization trial of Thaumatin performed with the same TbXo4 batch used with the first three assays. It should be noted that Thaumatin lot is a mixed of Thaumatin I, which we crystallize, and other isoforms of Thaumatin and also with traces of sweet proteins. Hence, the percentage of impurities of the lot may thus affect the crystallization output.

a)



b)



c)

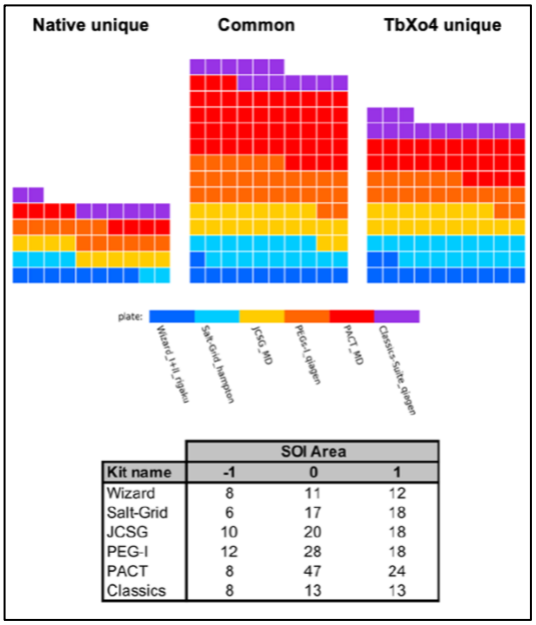
		Crystallization assays	
Protein lots		Native	TbXo4 Batch 2
Sigma-Aldrich	Lot 1	1, 2 & 3	1, 2 & 3
	Lot 2	4	4

Figure 3-5 *Number of crystallization hits of Thaumatin in absence (blue) and in presence (orange) of TbXo4, using 1+1 crystallization protocol test, the average(avg) and its standard deviation(error bars in red) and cumulative number (cml) of all the assays. c) Origin of protein lot and TbXo4 batch for each crystallization assay.*

To conclude, the variability of the number of crystallization hits from the crystallization assays of the 3 commercial proteins can be explained by the stochastic nature of the crystallization process, as well as the conditions of storage of the crystallization kits before the experiments which have an impact on the screens buffer species and their pH (Wilson et al., 2020).

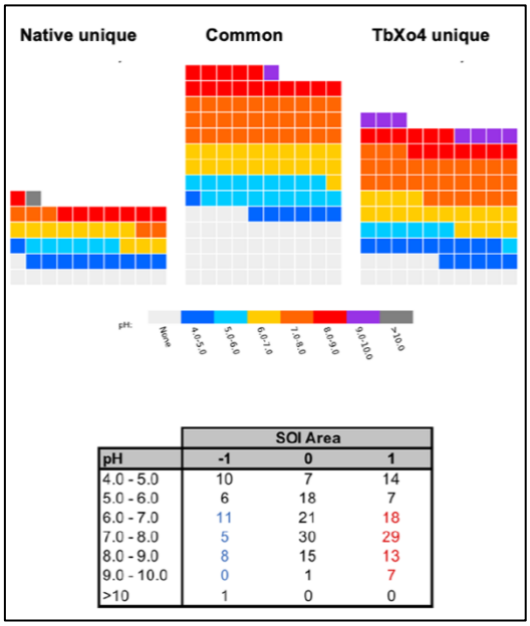
To complete, we have done a preliminary study of the correlation between the influence of TbXo4 and the composition of the crystallization screens (Jiang et al., 2020) in order to observe the effect of some parameters, such as the pH, on TbXo4 efficiency. In this paper, we exploited the crystallization outputs of 6 proteins (over 8) used in (Engilberge et al., 2017), including the three commercial ones. The results of the crystallization assays of the 6 proteins were exploited under a representation named as Subset Of Interest (SOI) (Figure 3-6). In this representation, each crystallization condition leading to a hit is represented by a square.

a)



	SOI Area		
Kit name	-1	0	1
Wizard	8	11	12
Salt-Grid	6	17	18
JCSG	10	20	18
PEG-I	12	28	18
PACT	8	47	24
Classics	8	13	13

b)



	SOI Area		
pH	-1	0	1
4.0 - 5.0	10	7	14
5.0 - 6.0	6	18	7
6.0 - 7.0	11	21	18
7.0 - 8.0	5	30	29
8.0 - 9.0	8	15	13
9.0 - 10.0	0	1	7
>10	1	0	0

Figure 3-6 *Representation and Subset Of Interest analysis (SOI Area tables) of the crystallization assays of six proteins against (a) commercial crystallization kit and (b) pH. Each square represents a crystallization hit. The three areas of SOI correspond to the unique hits observed for the proteins in the absence/presence of TbXo4 (area -1/+1, respectively) and observed in both cases (area 0). The pH value shown corresponds to the value presented in the crystallization conditions and available in the description of the commercial kits. None: indicates that no value is provided in the description (Jiang et al., 2020).*

The analysis carried out on the six proteins has shown a good synergy between TbXo4 and some crystallization kits used. In presence of TbXo4, the number of crystallization hits increases three times more in comparison to the conditions in absence of TbXo4 for both kits Salt-Grid and PACT (6 and 8 in absence of TbXo4; 18 and 24 in presence of TbXo4 respectively) (Figure 3-6 a). Thus, these two kits can be considered as starting point in order to limit, if needed, the number of crystallization conditions to be evaluated with TbXo4. As for the pH, the analysis of SOI representation of the 6 proteins in function of pH value has shown that TbXo4 promotes the appearance of crystals under crystallization conditions with a pH between 6 and 10 (Figure 3-6 b).

Such pH values may ensure that the carboxylate moiety of both aspartate and glutamate, which interact with TbXo4 (as mentioned in the Chapter 1.5.1) are completely deprotonated to facilitate a direct interaction with the Terbium ion(+III) (Jiang et al., 2020). Moreover, future studies will focus on the influence of each of the crystallisation components such as the buffer species, the precipitating agents (PEGs, salts...) or the presence of ions (Cations ex: Ca^{2+}, Mg^{2+} or anions) in the crystallisation

media. Furthermore, as we mentioned above, protein lots and their purification rate have a major effect on the nucleation and crystal growth of the protein. Also noted that the effectiveness of TbXo4 is protein dependent. Hence, further studies on the correlation between the nucleation properties of TbXo4 and the characteristics of the proteins such as their isoelectric point (pI) (Table 2-1), the proportion of acidic and basic residues or their ratio would be interesting.

In order to better characterize the crystallization space of a set of identical HTX experiments, for each protein (native and TbXo4), we have defined a cumulative results (cml) for each set of crystallization assays. For a set of identical crystallization assays of a protein, the cumulative results (cml) is defined by the sum of the number of crystallization hits obtained at least once (Figure 3-7). Thus, the cumulative value include all hits collected from several crystallization assays of a protein. This value will be used as a reference for the evaluation of TbXo4 variants, thereafter.

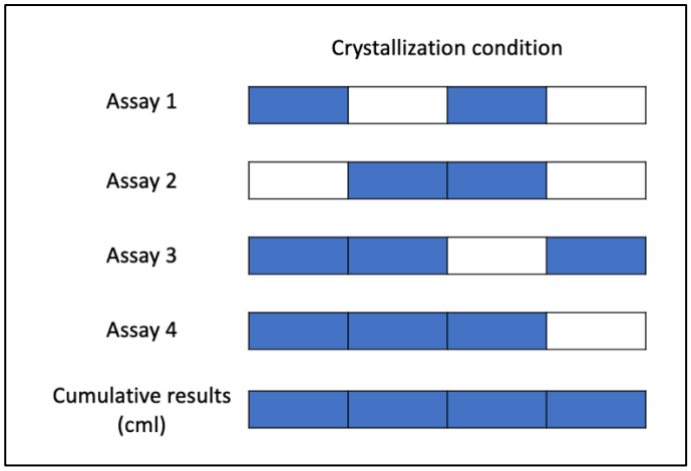


Figure 3-7 *Illustration of the cumulative crystallization results from four crystallization identical assays for the same protein. Assays 1&2 and 3&4 resulted in 2 and 3 hits, respectively. The corresponding cumulative value is 4.*

Furthermore, in order to better understand the behaviour of the new TbXo4 batches, we investigated the different possibilities that may affect the nucleating properties of TbXo4 such as the purity and composition of the used batch. As addressed at the beginning of this section, the new TbXo4 batches resulted from a modified and improved ligand synthesis protocol as a precursor initially used was no longer commercially available. The improved protocol, including also a new purification

approach, led to a better controlled and characterized final product. As mentioned earlier, the improved protocol happened to change the nature of the counter-ion from a carbonate HCO_3^- to a chloride. This change may have been considered as minor but not, as shown at the beginning (Figure 3-1 b) and as described hereafter.

Thus, to confirm the influence of the counter-anion on TbXo4 nucleating properties, I first performed HTX crystallization trials of the 3 commercial proteins in the presence of 10 mM TbXo4 and 10 mM of sodium bicarbonate ($NaHCO_3$ or HCO_3^-)(i.e one equivalent) using a 1+1 crystallization protocol. The results of this crystallization trials are shown in Figure 3-8 a (orange), with for comparison the average results of these 3 commercial proteins obtained above (TbXo4-avg in blue) and the results from Engilberge et al. (2017) (TbXo4-SE in grey), which were both performed using the 1+1 protocol. By comparing the average number of hits of the three commercial proteins, in presence of the new batch of TbXo4 (TbXo4-avg), to TbXo4-SE and TbXo4-HCO_3^-, we observe (Figure 3-8 a) that the addition of HCO_3^- has different influence depending on the protein. For HEWL, while the new batch shows a higher nucleating effect, both TbXo4-SE and TbXo4-HCO_3^- shows similar nucleating results with HEWL. So, by adding HCO_3^- to HEWL with the new batch of TbXo4, we found that, overall, TbXo4-HCO_3^- behaves like TbXo4-SE. As for ProteinaseK and Thaumatin, although the new batch is certainly more efficient than TbXo4-SE, however we do not observe a significant decrease of the nucleating effect of TbXo4 after adding HCO_3^-. This maybe be due to the fact that, for these two proteins, no direct coordination of TbXo4 with acidic residues occurs. Hence, we can propose that HCO_3^- has more effect on the nucleation properties of TbXo4, when direct coordination with aspartate and glutamate residues is involved. Indeed, the potential direct coordination of HCO_3^- with TbXo4 is likely to decrease its nucleating properties, at least for HEWL. However, the effect of the carbonate shall rather be considered as a modulator of TbXo4 nucleating effect, as it also avoids protein precipitation that can be considered as an “extreme” nucleating effect (Figure 3-1 b).

Additionally, to avoid protein precipitation at the mixing step, as seen in Figure 3-1 a, we have tested an alternative protocol (1+1+1 protocol) consisting in 1 volume of protein solution + 1 volume of 10 mM TbXo4 solution (dissolved in distilled water or protein buffer) + 1 volume of crystallization condition, on HEWL and two other proteins (Figure 3-8 b, in blue). For this protocol, the order of solution addition has to be

respected as different results are observed if for example, the addition of TbXo4 is the last step. We also evaluated the addition of 1eq. of HCO_3^- in the TbXo4 solution (Figure 3-8 b, in orange).

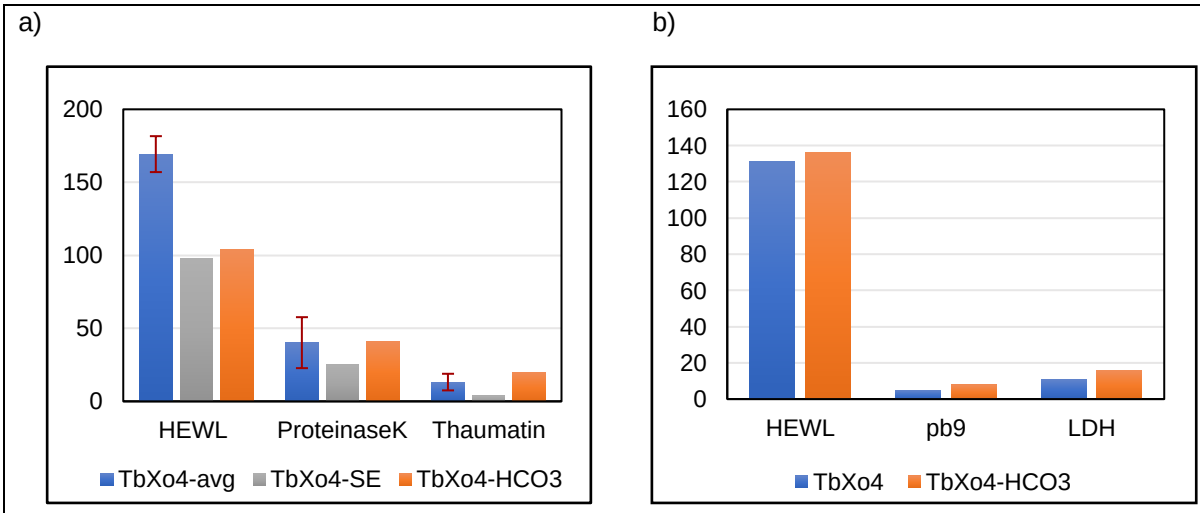


Figure 3-8 *Number of crystallization hits of a) three commercial proteins, HEWL (Hen Egg-White Lysozyme), ProteinaseK and Thaumatin in presence of TbXo4 batches, using 1+1 crystallization protocol test. In blue the average (avg) number of crystallization hits using the new TbXo4 batch and its standard deviation in red error bars, in grey the number from TbXo4-SE batch (Engilberge et al., 2017) and in orange the number in presence of 1eq. HCO_3^-. b) HEWL, pb9 & LDH, using 1+1+1 crystallization protocol. In blue the number of crystallization hits in absence of HCO_3^- and in orange the number in presence of 1eq. HCO_3^-.*

For HEWL, the number of hits using the 1+1+1 protocol seems closer to the value obtained with the 1+1 protocol in the presence of carbonate, indicating that the proposed protocol allows to control the extreme nucleating effect of TbXo4. This is confirmed by the two other proteins. Indeed, pb9 (Bacteriophage T5 distal tail tube protein) and the lactate dehydrogenase (LDH) were two clear examples of precipitation when the mixing was based on the TbXo4 powder (Figure 3-1 a). No significant differences between the absence and the presence of HCO_3^- are observed when this protocol is used. It is interesting to note that in the 1+1+1 protocol, the modulation of the nucleating effect is related to the fact that TbXo4 addition is performed in solution (TbXo4 powder dissolved in milliQ water) and may be related to a different kinetic of interaction with the protein surface.

In conclusion, it can be suggested that the 1+1+1 protocol should be favoured in order to make the most of the nucleating properties of TbXo4 without the risk of protein precipitation. If no exploitable hits are obtained, one can decide to test again the effect

of TbXo4 through direct mixing of the protein solution with the powder. Another reason in favour of the 1+1+1 protocol, is that 1+1+1 protocol gives similar results with and without HCO_3^- and that the use of HCO_3^- alters the pH of the final solution so that it may be deleterious to the protein of interest, hence it is more advisable to use 1+1+1 protocol without HCO_3^-.

For the following results, I have used TbXo4 without adding the bicarbonate, with 1+1 protocol for all my HTX crystallization assays and 1+1+1 protocol for manual crystallization plates, as it took several months before we have realized the issues with the new batch and have understood it. Moreover, it took also few other months before having a correct implementation of the 1+1+1 protocol at the high throughput crystallization platform (HTX-Lab). At that time I had already performed a large part of my assays.

3.2 TbXo4 variant description

As indicated in the introductory part of this chapter, we have decided to exploit the chemistry offered by the Xo4 scaffold to study the effect of different moieties on the nucleating properties. Our collaborators at the ENS-Lyon have designed and synthetized several variants (Figure 3-9) that can be divided into three main families.

Variants where the terbium ion has been changed for another lanthanide or metal (Figure 3-9 a). Indeed, the ionic radius is different through the whole lanthanide series. This may have an influence on the properties of the complex like the dynamic and the distribution of charges within the molecule. For that, variants encompassing atoms with a higher ionic radius (LaXo4 and EuXo4) or with a lower ionic radius (YbXo4 and LuXo4) were tested. Also, the change of the Terbium atom to a Bismuth atom, the most stable atom in the periodic table with an ionic radius value between Europium and Terbium ones, was evaluated.

Variants with modifications on the picolinate moiety such as addition of sulfate moiety or iodine atoms (Figure 3-9 b). This family of variants were developed to modulate the hydrophobicity and the overall electrostatic charge of the resulting complex.

Variants exploiting the free nitrogen position present in the triazacyclononane (TACN) macrocycle with addition of a propanol function (TbXo4-C_3H_6-OH or TbXo4-OH), a propane-sulfonate function (TbXo4-C_3H_6-SO_3^- or TbXo4-SO_3^-), a 1-dimethylamino

propane function (TbXo4-C_3H_6-$NMet_2$ or TbXo4-$NMet_2$) or polyethylene glycol extension (TbXo4-OPeg and TbXo4-Peg5OH) (Figure 3-9 c). This family of variants were developed to modulate the steric hindrance and/or the overall charge.

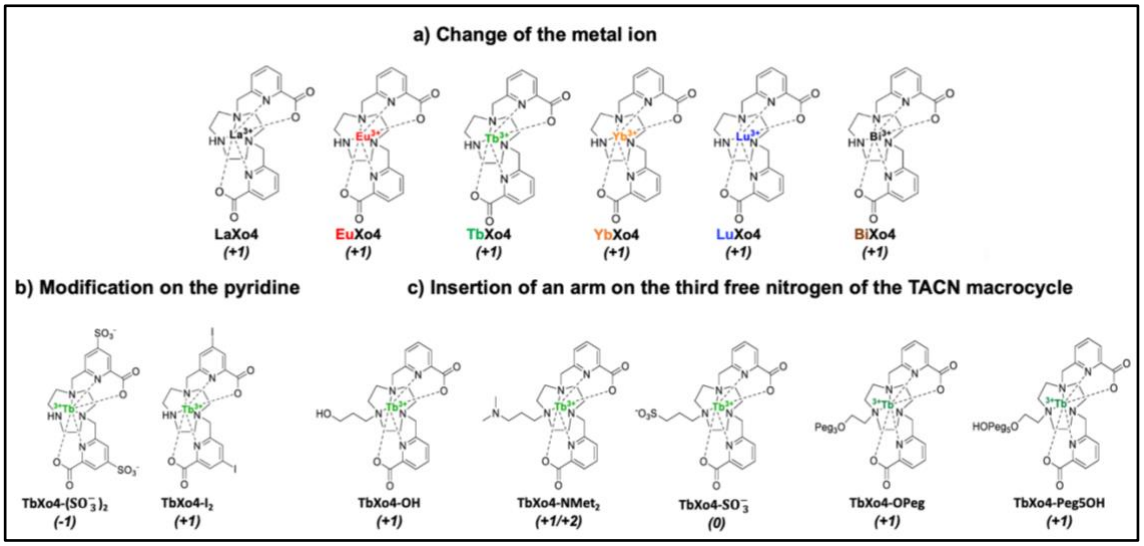


Figure 3-9 *Crystallophore variants produced by the team of Olivier Maury at ENS-Lyon. The three families are described within the text. The values between brackets indicate the electrostatic charge of the complex.*

3.2.1 Evaluation of variant nucleating properties

To compare the nucleating properties of the different variants with those of TbXo4, we applied the same methodology used to highlight TbXo4 nucleating properties (Engilberge et al., 2017). A comparative experiment was performed by comparing the crystallization output in the absence (native) and in the presence of each variant. We have evaluated these complexes on the three commercial proteins (HEWL, ProteinaseK, Thaumatin), by screening 576 crystallization conditions from various commercial kits (Material & methods, Table 2-2) at the high-throughput crystallization platform (HTX-Lab, EMBL Grenoble). Protein solutions at 20 mg/mL were used to dissolve each variant to a final concentration of 10 mM.

The results of these comparative experiments are shown in the histograms depicted in Figures 3-10 to 3-15. It should be noted that the results of LuXo4 and TbXo4-OH are the cumulated results from several crystallization assays of these three commercial proteins, in addition of TbXo4-$NMet_2$ result with ProteinaseK.

The choice of the most interesting variants depends on several criteria. The variant should: 1) cover most of the protein's native hits as shown for TbXo4, 2) induce more unique hits comparing to TbXo4 when considering each model protein, 3) and then cover TbXo4 hits or show a good complementarity when considering the results of the three model proteins. Moreover, to facilitate the choice, I have first compared the complexes within each variant family to select the most promising variant of each family, then I have chosen the most efficient one from the three selected by comparing them with the results from TbXo4 cumulated crystallization results.

3.2.2 Presence of self-crystallization

In a recent study (Roux et al., 2021), we have shown that some lanthanide complexes may induce false-positive results. As opposed to heterogeneous nucleating agents which induce false-positive by facilitating salt crystallization in conditions that contains significant levels of salt (Khurshid et al., 2014), lanthanide complexes can self-crystallized in limited conditions found in the commercial crystallization kits. This is particularly true with the tris-dipicolinate molecule $Na_3[Eu(DPA)_3]$. To highlight this self-crystallization, the molecule was dissolved in water and considered as a sample for HTX run with the same 576 crystallization conditions used to evaluate the nucleating properties of the crystallophore. In such conditions, the spontaneous crystallization of $Na_3[Eu(DPA)_3]$ was observed in 10% of the conditions at a molecule concentration of 25 mM, and this ratio increased up to 30% at 100 mM $Na_3[Eu(DPA)_3]$. In the same study we also evaluated TbXo4 and observed no spontaneous crystallization at 10 mM concentration and only 3 over 576 conditions at 100 mM concentration. Obviously, I did the same experiment for each variant at a concentration of 10 mM. Only TbXo4-$(SO_3)_2$ showed spontaneous crystallization with 12 over 576 conditions that contains 200 mM of calcium salt (calcium chloride and calcium acetate).

3.2.3 Results with Hen Egg-White Lysozyme

For the first model protein, HEWL, the number of crystallization hits obtained after screening assays with and without the variants are presented in Figure 3-10.

Within the first variant family, LuXo4 is the most interesting as it induces the highest number of crystallization hits and as it covers the largest portion (78%) of HEWL native hits. It also shows the highest number of unique hits.

For the second variant family, between the two variants present in this family, TbXo4-$(SO_3)_2$ is the variant that covers most native hits and induces more unique hits of the protein model, similarly to LuXo4 results. Moreover, the second variant (TbXo4-I_2) has a poor solubility in the presence of the proteins leading to lots of conditions with precipitates. Hence, TbXo4-$(SO_3)_2$ can be considered the most interesting variant in this family. However, as mentioned in 3.2.2, TbXo4-$(SO_3)_2$ is also the only variant that self-crystallizes.

Concerning the third family of variants, we can clearly see that TbXo4-OH is the most competent variant as it covers 85% of HEWL native hits and it possesses the highest nucleating effect compared to other variants (Figure 3-10).

Figure 3-11 presents the comparison of HEWL crystallization outputs of the variants with those obtained with TbXo4. Globally, we observe the same trends as described previously. We found that TbXo4-OH covers a major portion (88%) of TbXo4 hits compared to LuXo4 which cover 76% of TbXo4 hits. Also,TbXo4-OH induces more unique crystallization hits comparing to TbXo4.

Overall, TbXo4-OH is the most efficient variant in the generation of crystallization hits for HEWL.

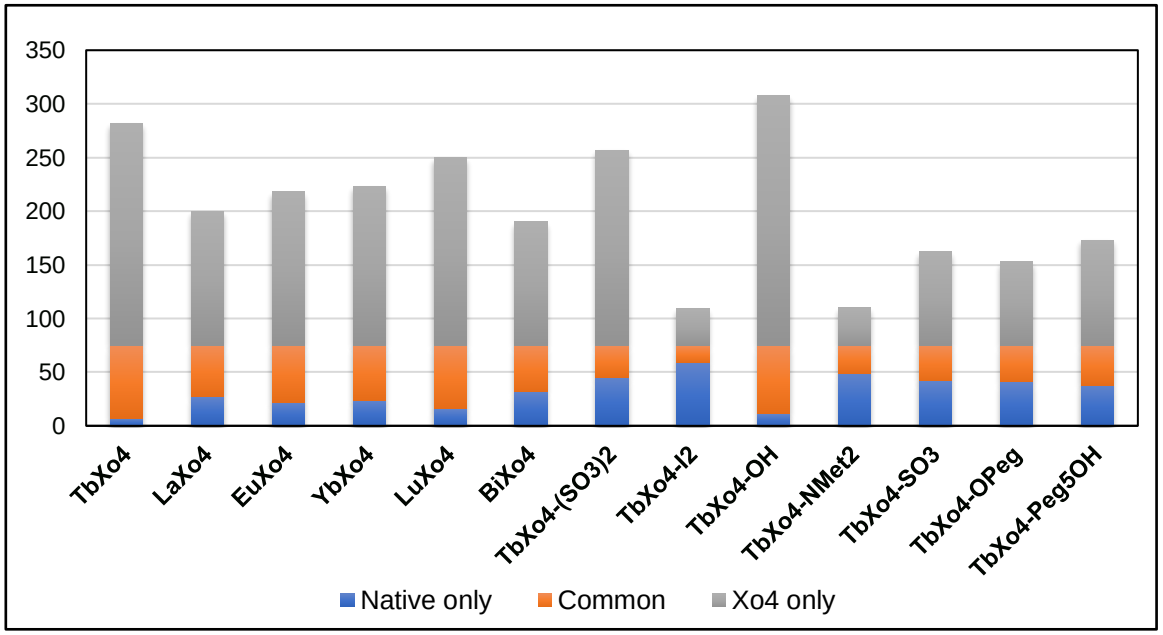


Figure 3-10 *Results of the automated crystallization screening of crystallophore variants with HEWL. Drops observation after 30 days: number of unique hits for native protein in blue, common hits (native and Xo4) in orange, Xo4 variants unique hits in grey.*

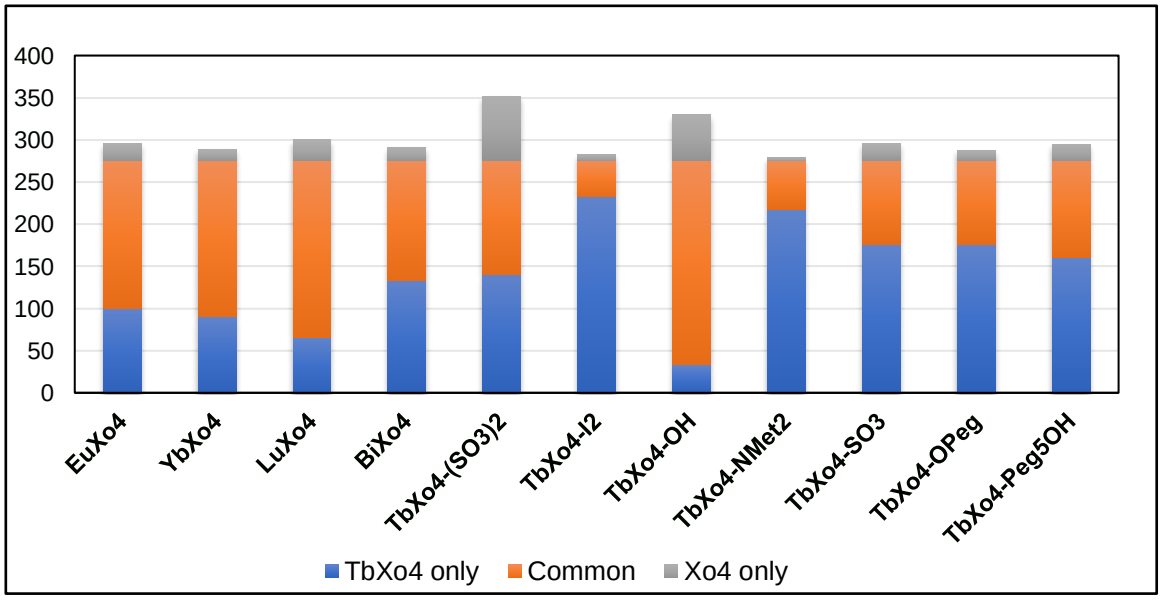


Figure 3-11 *Comparison of variant versus TbXo4 crystallization outputs in the case of HEWL. Drop observation after 30 days: TbXo4 unique hits in blue, common hits (TbXo4 and Xo4 variants) in orange, Xo4 variants unique hits in grey.*

3.2.4 Results with ProteinaseK

The results obtained for the second protein model, ProteinaseK, are depicted in Figure 3-12. While almost all variants covers around 30% of the native hits, three variants, namely TbXo4-OH, TbXo4-I2, and TbXo4-NMet2 induces a high number of unique hits, higher than the number of hits obtained with TbXo4 cumulated results. As for HEWL, the comparison with TbXo4 shows again that TbXo4-OH covers a large portion of TbXo4 hits and induces more unique hits (Figure 3-13). Hence, we can again consider that TbXo4-OH variant is the most efficient variant. However, unlike in the case of HEWL, the two other variants TbXo4-I2, and TbXo4-NMet2 present promising outputs in the case of ProteinaseK.

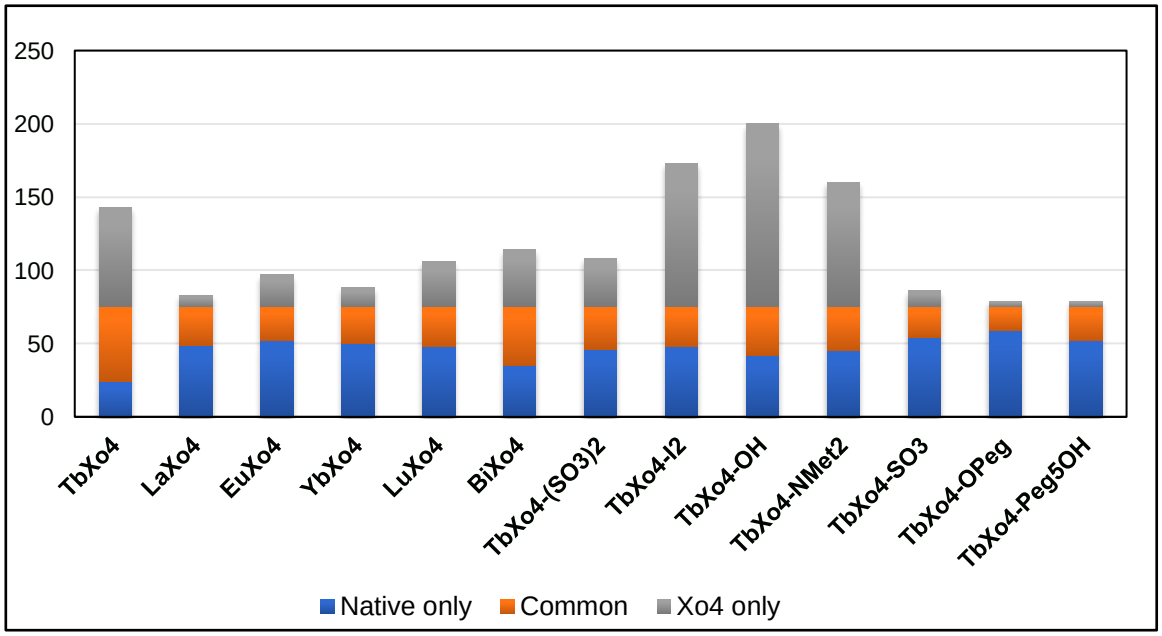


Figure 3-12 *Results of the automated crystallization screening of crystallophore variants with ProteinaseK. Drop observation after 30 days: number of unique hits for native protein in blue, common hits (native and Xo4) in orange, Xo4 variants unique hits in grey.*

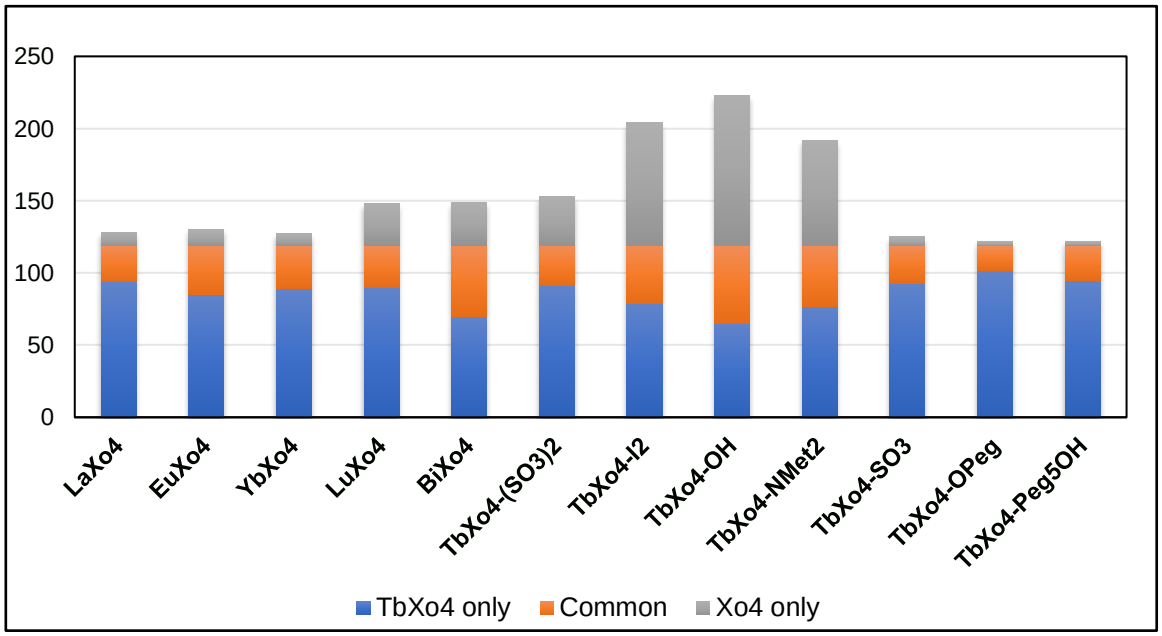


Figure 3-13 *Comparison of variant versus TbXo4 crystallization output in the case of ProteinaseK. Drop observation after 30 days: TbXo4 unique hits in blue, common hits (TbXo4 and Xo4 variants) in orange, Xo4 variants unique hits in grey.*

3.2.5 Results with Thaumatin

Concerning the third model protein, Thaumatin, the HTX screening carried out leads to a slightly different result from those observed for the first two proteins (Figure 3-14). In particular and unexpectedly, the LaXo4 variant produced the highest number of unique hits, having an important overlap with the native hits. However, this result is to be underestimated as this molecule has a lower solubility than the other variants. As with ProteinaseK, the TbXo4-I2 variant also has a large number of unique hits. However, it covers few of the native hits. The two variants, LuXo4 and TbXo4-(SO3)2, show an interesting number of unique hits and a good overlap with the native hits. Finally, as for the other two proteins, the TbXo4-OH variant has a good overlap with the native hits and a large number of unique hits.

The comparison with TbXo4 confirms again that TbXo4-OH covers a large portion of TbXo4 hits and induces more unique hits (Figure 3-15) and is the most interesting variant. It should be noticed that TbXo4-I2 variant present an almost perfect complementarity with TbXo4 as it presents a small number of common hits.

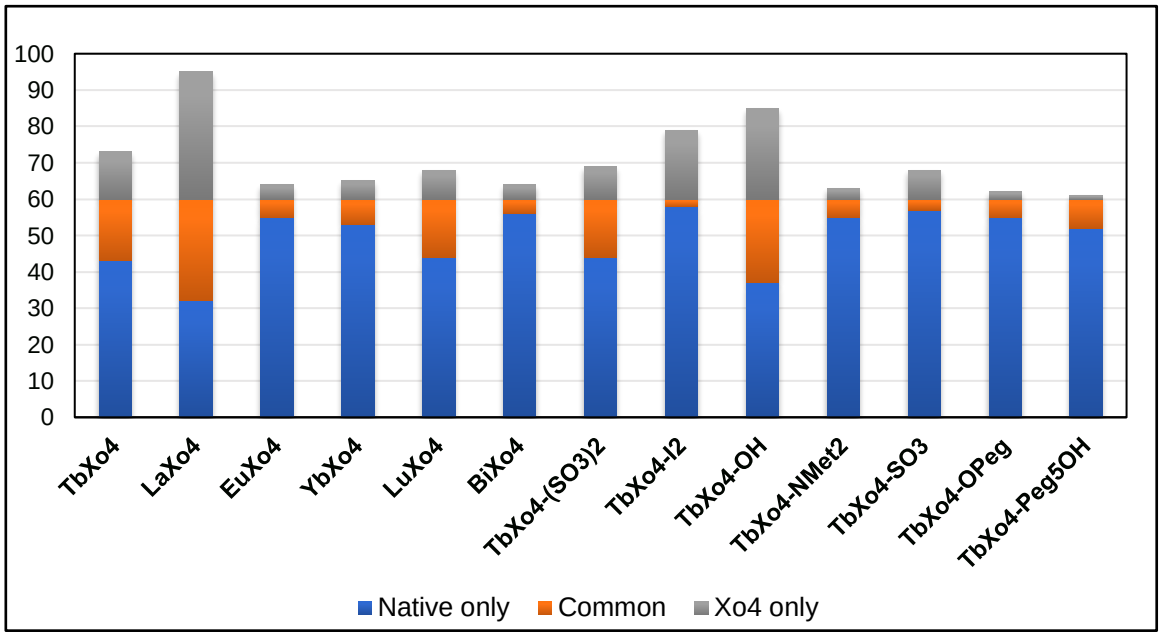


Figure 3-14 *Results of the automated crystallization screening of crystallophore variants with Thaumatin after 30 days: number of unique hits for native protein in blue, common hits (native and Xo4) in orange, Xo4 variants unique hits in grey.*

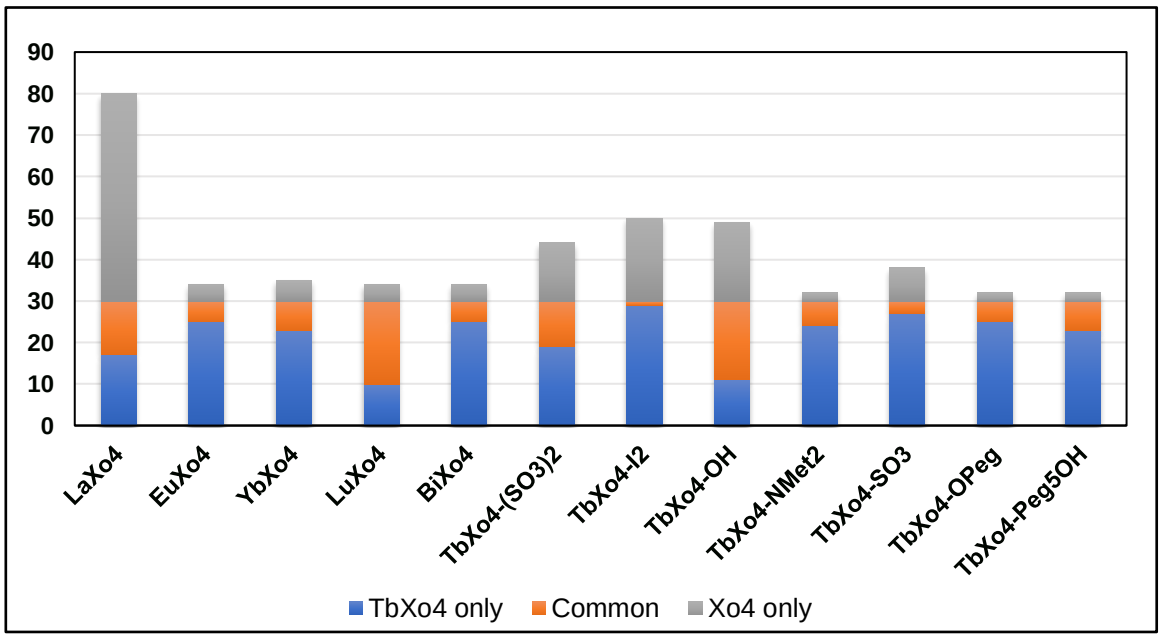


Figure 3-15 *Comparison of variant versus TbXo4 crystallization outputs in the case of Thaumatin. Drop observation after 30 days: TbXo4 unique hits in blue, common hits (TbXo4 and Xo4 variants) in orange, Xo4 variants unique hits in grey.*

3.2.6 Conclusion on Xo4 variant nucleating properties.

The evaluation of the nucleating properties of the different variants shows contrasted results. It was not anticipated that the metal may have an importance on the nucleating properties of our molecules and should be considered in the future. Despite the fact that the ionic radius may influence the overall molecule dynamic, this observation remains misunderstood. The results for Family2 of variants are also contrasted. It is difficult to say if one of the two molecules is more interesting because their efficiency depends on the protein tested. The situation of the Family3 is clearer. Indeed, the variants based on PEG-like extensions are of very limited interest. On the other hand, as we have seen with the three model proteins, the TbXo4-OH variant is the one that leads to interesting crystallization results, which makes it the most efficient molecule in addition to the parent molecule, TbXo4. As for the family2, the two other variants present an interest which remains protein dependent.
All together, these results and conclusion would benefit from complementary experiments using more real-life proteins.

3.3 Determination of TbXo4 and Xo4 variants binding mode in nucleating condition

In order to understand the interaction and the binding mode of TbXo4-OH and possibly, get insight on its nucleating properties as well as of its parent molecule TbXo4, we have collected diffraction data on crystals of HEWL co-crystallized with 10 mM of TbXo4-OH. To facilitate the analysis, we also collected data on TbXo4-SO3 and TbXo4-NMet2 HEWL derivatives. Crystals were grown in the conventional HEWL crystallization condition that contains NaCl with sodium acetate buffer at pH=4.5 (Table 3-1). Data collection were performed at L_{III} absorption edge of Terbium for all the crystals to unambiguously position the terbium atom.
Data quality and refinement statistics are given in Table 3-2 and 3-3, respectively.

Table 3-1 *Crystallization conditions of HEWL with Xo4 variants.*

Protein	Crystallization condition	Cryoprotectant
HEWL with 10 mM TbXo4	800 mM NaCl, 100mM NaAc pH=4.6	1 M NaCl 100mM NaAc pH=4.6, 30% PEG400
HEWL with 10 mM TbXo4-OH	950 mM NaCl, 100mM NaAc pH=4.6	1 M NaCl 100mM NaAc pH=4.6, 30% PEG400
HEWL with 10 mM TbXo4-SO3	950 mM NaCl, 100mM NaAc pH=4.6	1 M NaCl 100mM NaAc pH=4.6, 30% PEG400
HEWL with 10 mM TbXo4-NMet2	950 mM NaCl, 100mM NaAc pH=4.6	1 M NaCl 100mM NaAc pH=4.6, 30% PEG400

During his PhD thesis, Sylvain Engilberge determined the structure of HEWL in the presence of 100 mM TbXo4 (Engilberge et al., 2018) (PDB code: 6F2I). However, in this thesis, as we would like to understand the mechanism of nucleation of TbXo4, I choose to study the interaction of HEWL with TbXo4 in its nucleation condition, therefore with 10 mM of TbXo4. Hence, to limit bias in structure comparisons, I produced HEWL crystals in the presence of 10 mM TbXo4 and determined the corresponding structure.

Data quality and refinement statistics of this structure are given in Table 3-2.

3.3.1 Structure of HEWL with 10 mM TbXo4

As mentioned in the Chapter 1.5.4, HEWL structure obtained with 100mM TbXo4 shows the presence of 2 TbXo4 sites with occupancies of 0.8 and 0.6 (B-factor: 18.92 and 17.54 $Å^2$), respectively. These two TbXo4 were found associated as a dimer in the asymmetric unit, where the carbonyl unit of the picolinate arm from the first crystallophore molecule coordinates the metal of the second lanthanide complex (Figure 1-15 b). Also, the first complex coordinates with the carboxy oxygens of Aspartate 101 (D101) of the protein and it forms an hydrophobic interaction with tryptophan 62 (W62). In addition, the ligand of the second crystallophore form an hydrophobic interaction with tryptophan 123 (W123') of a symmetrical molecule. Also, this second TbXo4 forms a ionic interaction between the carbonyl group of one

picolinate moiety and the ammonium one of lysine 33 (K33) of a symmetrical protein (Engilberge et al., 2018).

Table 3-2 *Structural data collection and refinement statistics of HEWL in the presence TbXo4.*

	HEWL 10 mM TbXo4
Data collection	
Synchrotron-Beamline	SOLEIL-Proxima1
Wavelength (Å)	1.648
Space group	P 4_3 2_1 2
Resolution (Å) [a]	37.49 - 1.70 (1.79 - 1.70)
Cell dimensions	
a, b, c (Å)	77.474, 77.474, 37.491
R_{sym} or R_{merge} (%) [a]	9.7 (33.6)
R_{pim}(%) [a]	2.6 (7.3)
$CC_{1/2}$ (%) [a]	99.8 (98.6)
I/σ(I) [a]	22.8 (9)
Completeness (%) [a]	99.7 (98.8)
Redundancy [a]	24.1 (23.5)
Number of unique reflections [a]	13059 (1848)
Refinement	
Resolution (Å)	33.75 - 1.70
Number of reflections	13015 (1252)
R_{work} / R_{free} [b] (%)	16.99 / 20.96
Number of atoms	
Protein	1019
Ligands/ion	61
Solvent	133
Mean B-factors (Å^2)	32.33
Protein	30.90
Ligand/ion	50.30
Water	38.19
Ramachandran plot	
Favoured regions (%)	98.43
Allowed regions (%)	1.57
Outlier regions (%)	0.0
R.m.s. deviations	
Bond lengths (Å)	0.017
Bond angles (°)	1.49

[a] Values in parentheses are for highest-resolution shell. [b] R_{free} was calculated as the R_{work} for 5% of the reflections that were not included in the refinement. [c] rmsd. root mean square deviation.

For HEWL structure co-crystallized with 10 mM TbXo4, we found one principal site as confirmed by the anomalous Fourier map (Figure 3-16 a) with a lower occupation rate of 0.6 and a higher B-factor of 53.4 $Å^2$ than the one obtained in HEWL 100mM TbXo4 structure, leading to a less well defined electron density (Figs 3-16 c et 3-16 e). Also, according to the F_o-F_c map (Figure 3-17), it shows that there is residual density related to positive sigma values around the complex, showing that TbXo4 presents other conformations in this structure of HEWL. However, it was not possible to model these other conformations within the electron density.

We also found a very low second anomalous peak that can be related to the Tb ion from the second TbXo4 molecule as observed in the HEWL 100 mM TbXo4 structure. However, no clear density map related to the whole ligand is observed around this second lanthanide ion. Hence, we may conclude from the comparison between 10 and 100 mM TbXo4 structures of HEWL that the presence of the second ligand is required to lock the conformation of the first complex.

Despite these differences, we always observe a direct interaction of the terbium atom with Aspartate 101 (D101) and the hydrophobic interaction with tryptophane 62 (W62) (Figure 3-16 b-d-f).

To conclude, the 10 mM TbXo4 structure of HEWL shows us that under "nucleating" conditions, the binding site presents the already described mode of interaction even though it shows some dynamics. The co-crystallization at higher TbXo4 concentration leads to the formation of a stable TbXo4 dimer on the HEWL surface. It seems that it is the presence of the second molecule that leads to lock the molecule bound to D101 in a favoured conformation.

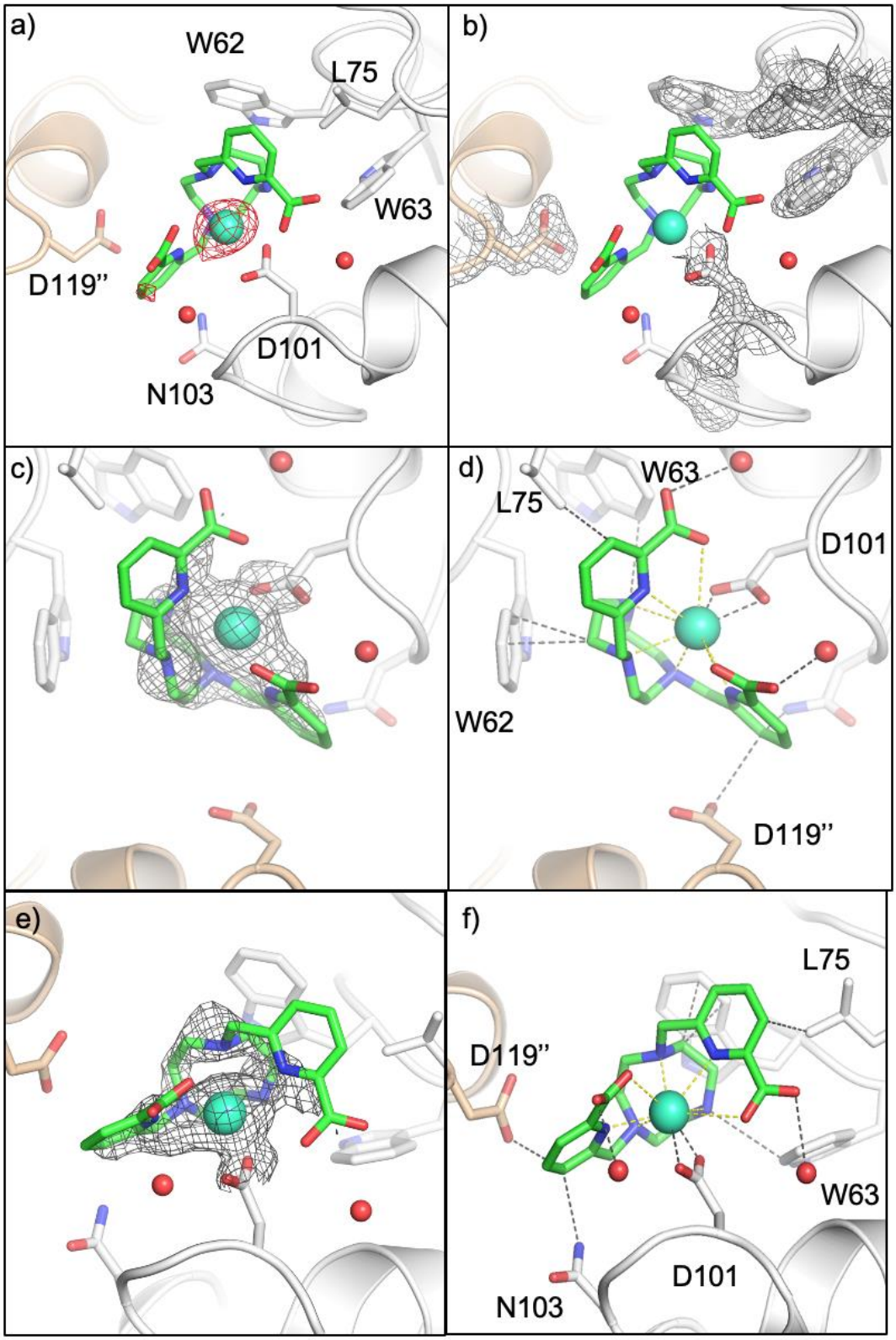


Figure 3-16 *Structure of HEWL co-crystallised with 10 mM TbXo4. a) anomalous Fourier map (contour: 5 sigma). 2Fo-Fc electron density map (contour: 1 sigma) of b) TbXo4 binding residues and c&e) the TbXo4 molecule (two views). d&f) Interaction sites of TbXo4 (Two same views as c&e panels).*

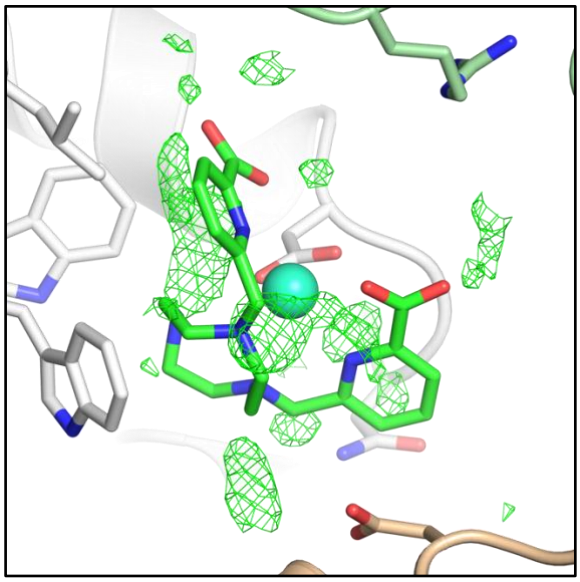


Figure 3-17 *Fo-Fc electron density map contoured at +3 sigma around the TbXo4 molecule in 10 mM crystallophore HEWL structure.*

3.3.2 HEWL structures with selected Family2 variants

As mentioned previously, I determined the HEWL structure in the presence of 10 mM TbXo4-OH and compared it to the ones of two other variants as well as the one in the presence of 10 mM TbXo4.

Data quality and refinement statistics are given in Table 3-3.

Before any binding description and as a first observation, the TbXo4-OH HEWL derivative present a better diffracting behaviour compared to the behaviour of TbXo4-SO3 and TbXo4-NMet2 derivatives, even of TbXo4. Indeed, despite the fact the data collection parameters were chosen to fully exploit the anomalous contribution in spite of the data resolution, the integration statistics indicate that this variant induced a better diffracting sample as illustrated by the I/sig(I) value at high resolution. This was later confirmed from refinement of the structure leading to lower overall B-factors (Table 3-3)**.**

It should be noted that we would have been able to collect data at a higher resolution than 1.7 Å if it was not for the detector distance limitation.

Table 3-3 *Structural data collection and refinement statistics of HEWL in the presence Xo4 variants.*

	HEWL 10 mM TbXo4-OH	HEWL 10 mM TbXo4-SO3	HEWL 10 mM TbXo4-NMet2
Data collection			
Synchrotron-Beamline	SOLEIL-PX1	SOLEIL-PX1	SOLEIL-PX1
Wavelength (Å)	1.648	1.648	1.648
Space group	P 4_3 2_1 2	P 4_3 2_1 2	P 4_3 2_1 2
Resolution (Å) [a]	38.57 - 1.69 (1.78 - 1.69)	38.7 - 1.74 (1.84 - 1.74)	38.76 - 1.72 (1.82 - 1.72)
Cell dimensions a, b, c (Å)	77.143, 77.143, 37.695	77.404 77.404 37.602	77.521 77.521 37.793
Rsym or Rmerge (%)[a]	6.6 (19.2)	7.4 (85.5)	8.3 (61.7)
Rpim(%)[a]	2.5 (5.4)	2.8 (18.4)	2.4 (13.8)
CC1/2 (%)[a]	99.9 (99.4)	99.8 (92.9)	99.8 (96.5)
I/σI [a]	34.3 (15.8)	24.8 (4.0)	24.2 (6.2)
Completeness (%)[a]	99.9 (99.4)	99.9 (99.2)	99.9 (99.4)
Redundancy [a]	24.3 (23.8)	24.1 (23.8)	24.2 (23.5)
Number of unique reflections [a]	13258 (1874)	12161 (1719)	12670 (1776)
Refinement			
Resolution (Å)	34.5 - 1.69	34.62 - 1.74	34.67 - 1.72
Number of reflections	13213 (1276)	12122 (1177)	12626 (1205)
Rwork / Rfree [b] (%)	15.64 / 19.48	19.73 / 23.33	17.41 / 19.04
Number of atoms			
Protein	1020	1009	1021
Ligands/ion	69	73	77
Solvent	154	118	131
Mean B-factors (Å2)	23.72	33.25	30.23
Protein	22.32	32.15	29.33
Ligand/ion	29.66	42.87	33.69
Water	31.43	39.04	36.13
Ramachandran plot			
Favored regions (%)	99.21	98.43	99.21
Allowed regions (%)	0.79	1.57	0.79
Outlier regions (%)	0.0	0.0	0.0
R.m.s. deviations			
Bond lengths (Å)	0.008	0.014	0.005
Bond angles (°)	1.18	1.01	0.78

[a] Values in parentheses are for highest-resolution shell. [b] R_{free} was calculated as the R_{work} for 5% of the reflections that were not included in the refinement.

[c] rmsd. root mean square deviation.

HEWL structures in the presence of 10 mM of each of the three variants show the presence of a main binding site involving the aspartate D101 as observed in the structure obtained with TbXo4. In addition to D101, another residue involved in the coordination of TbXo4, W62 was globally found as evidenced by the corresponding 2Fo-Fc electron density maps (Figures 3-18, 19, 20 b). The refinements lead to occupancy values (and thermal B factors) for terbium atoms of 0.75 (B-factor = 28.46 $Å^2$), 0.75 (B-factor = 40.68 $Å^2$) and 0.55 (B-factor = 40.06 $Å^2$) for the variants TbXo4-OH, TbXo4-SO3 and TbXo4-NMet2, respectively. These values are to be compared with the terbium atom values, 0.6 (B-factor= 53.42 $Å^2$), obtained for TbXo4. As with the quality of the data, the occupancies seem to indicate a better binding of the TbXo4-OH variant to the HEWL surface and on this basis the following scale can be proposed: TbXo4-OH > TbXo4-SO3 > TbXo4 > TbXo4-NMet2. This can also be seen in the quality of the 2Fo-Fc electron density map corresponding to each of the molecules (Figures 3-18, 19, 20: c & e).

Another observation goes in the same direction. Indeed, the anomalous Fourier synthesis (Figure 3-18, 19, 20 a) shows the presence of a weak peak in the case of the structures of the variants TbXo4-SO3 and TbXo4-NMet2, as well as TbXo4 (Figure 3-16 a). This indicate the weak presence of a second molecule, as the distance between the two anomalous peaks are quite close to the one observed in the HEWL structure in the presence of 100 mM TbXo4.

Regarding the extension present in each of the variants, it should be noted that it was not possible to model these extensions unambiguously, as the corresponding 2Fo-Fc electron density maps are weak (Figures 3-18, 19, 20 c). Hence, no interaction between the extension functions and the surrounding amino acids could be described. However, their presence must create steric hindrance which, combined with the direct interaction of terbium with D101 (which is the major binding point), induces a different conformation of the Xo4 ligand than that obtained for TbXo4 (Figure 3-21 a,b). This change leads to slightly shorter distances between the oxygens of the carboxylate function of D101 and the Tb atom: 2.7 Å in the case of TbXo4 and 2.6 Å for the variants. Moreover, the new conformation leads to the establishment of an interaction between arginine 5 of a symmetric molecule and the carboxylate of one of the two picolinates of the variants (Figures 3-18 d & f , 3-19 d, 3-20 f) with interaction distances of 3.0, 2.9 and 3.2 Å respectively for each variant.

To conclude, TbXo4-OH seems to bind better on the protein surface, perhaps due to the lower steric hindrance of the OH extension comparing to the other two variants. Hence, it would be interesting to study the variant interactions with HEWL in other crystallization conditions that may allow the description of Xo4 interactions.

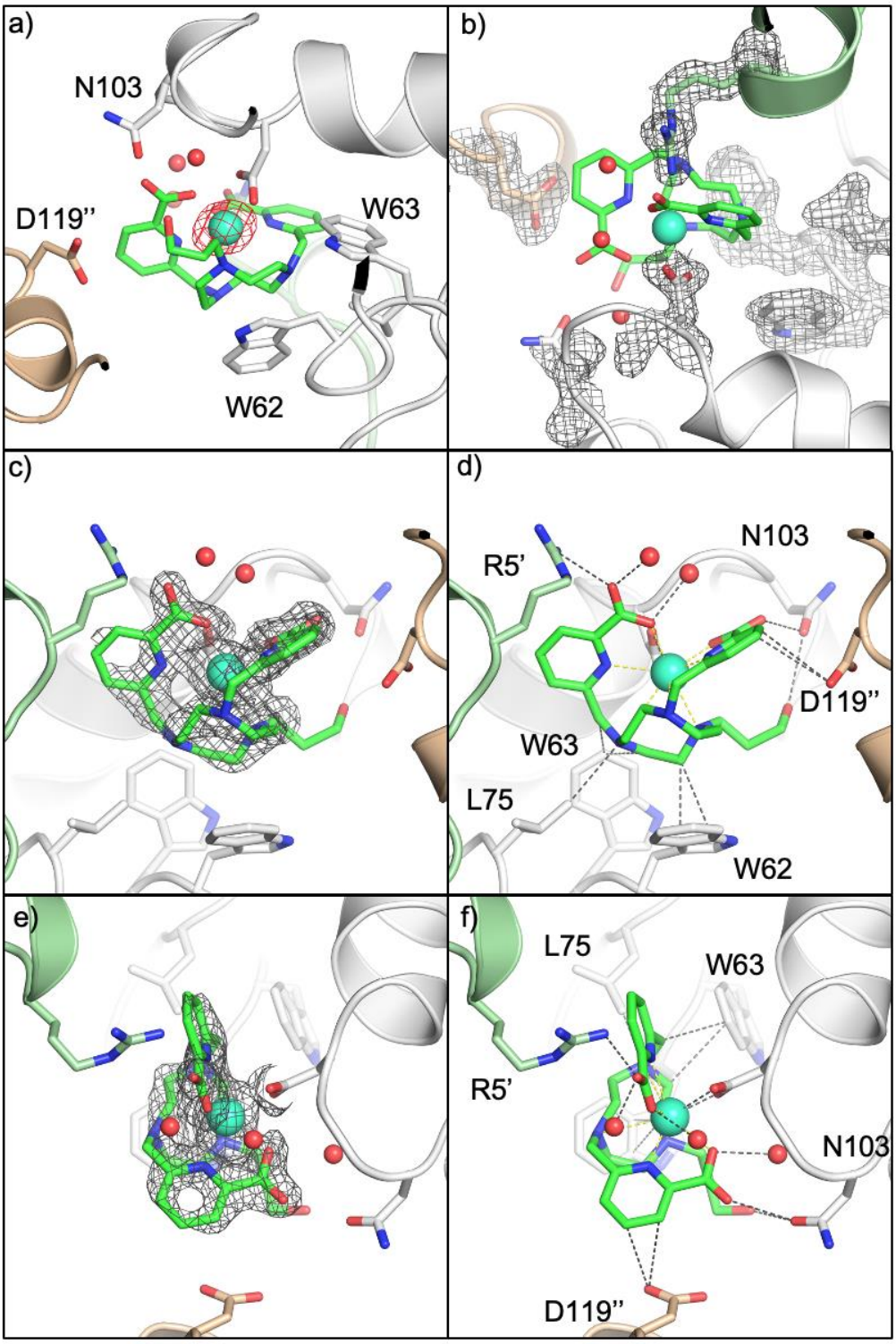


Figure 3-18 *Structure of HEWL co-crystallised with 10 mM TbXo4-OH. a) anomalous Fourier map (contour: 5 sigma). 2Fo-Fc electron density map (contour: 1 sigma) of b) TbXo4-OH binding residues and c&e) the TbXo4-OH molecule (two views). d&f) Interaction sites of TbXo4-OH (Two same views as c&e panels).*

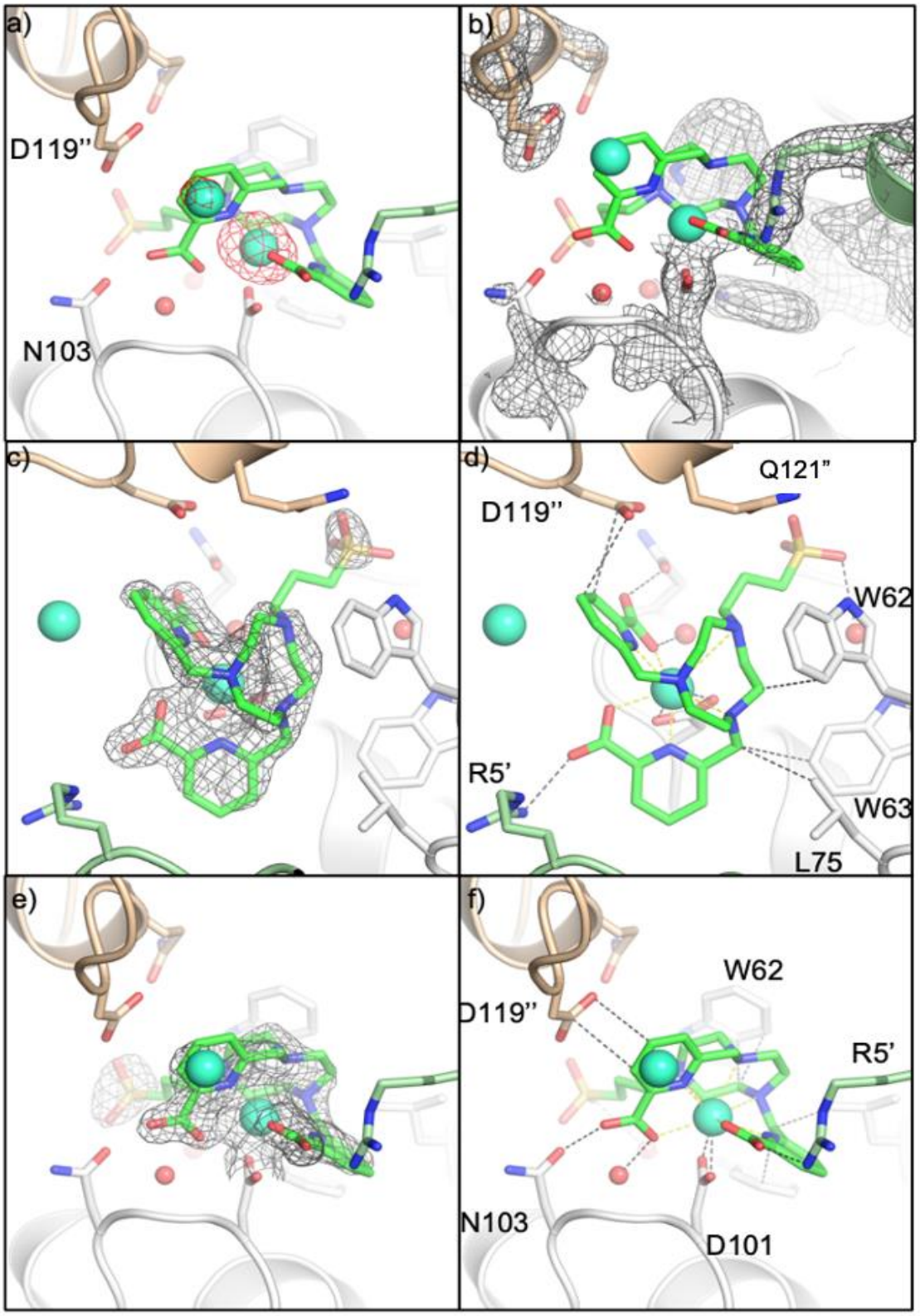


Figure 3-19 *Structure of HEWL co-crystallised with 10 mM TbXo4- SO3. a) anomalous Fourier map (contour: 5 sigma). 2Fo-Fc electron density map (contour: 1 sigma) of b) TbXo4- SO3 binding residues and c&e) the TbXo4- SO3 molecule (two views). d&f) Interaction sites of TbXo4- SO3 (Two same views as c&e panels).*

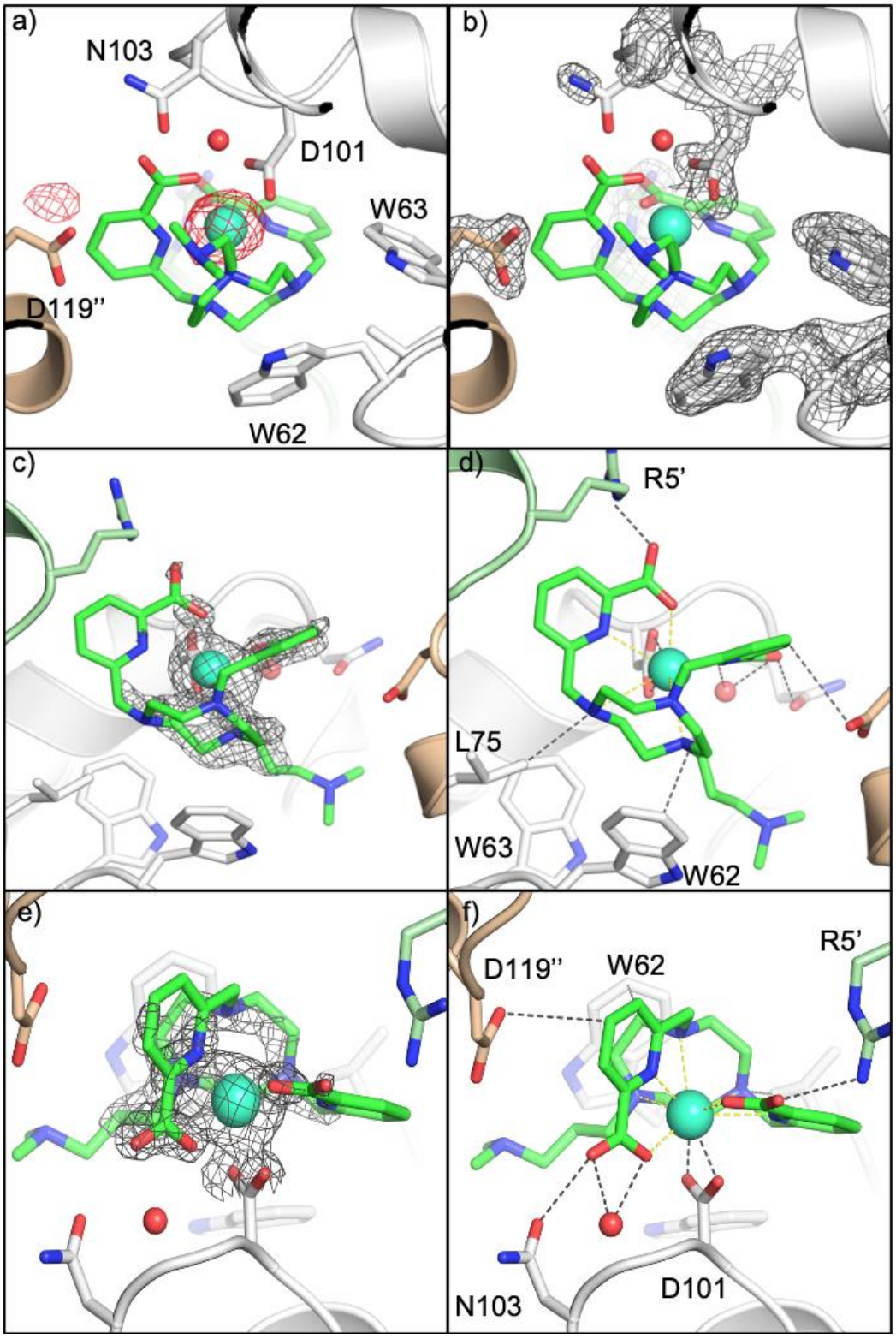


Figure 3-20 *Structure of HEWL co-crystallised with 10 mM TbXo4-NMet2. a) anomalous Fourier map (contour: 5 sigma). 2Fo-Fc electron density map (contour: 1 sigma) of b) TbXo4-NMet2 binding residues and c&e) the TbXo4-NMet2 molecule (two views). d&f) Interaction sites of TbXo4-NMet2 (Two same views as c&e panels).*

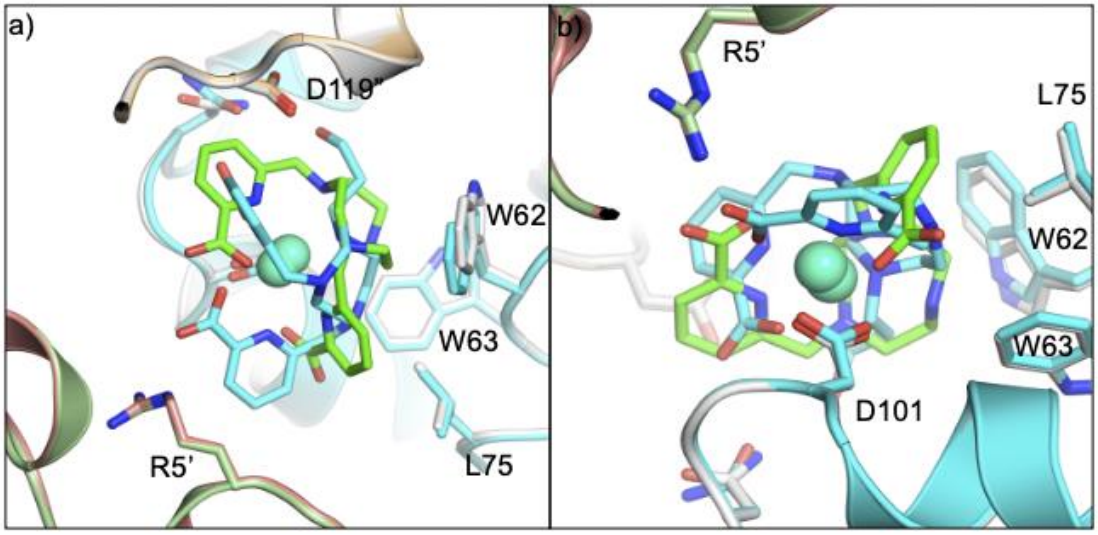


Figure 3-21 *Superimposition of HEWL structures with TbXo4 and TbXo4-OH (TbXo4 in green and TbXo4-OH in blue).*

3.3.3 QM/MM molecular simulations

Further analysis on the binding mode of TbXo4 and its three variants with HEWL were performed by All-atom molecular dynamic simulations to compare the experimental observation on TbXo4 and thus better understand the property of these new variants and the mechanism of their interaction. These analysis were performed by our collaborators at ENS-Lyon (E. Dumont & N. Gillet) for 200 ns in solution, in presence of the lanthanide complexes using the crystallographic structures as starting point.
As observed experimentally, the results of the simulations indicate that the main interaction of the complex with HEWL is similar for TbXo4 and the three variants and involves the direct coordination of the terbium atom with D101 and strong hydrophobic (CH-π) interactions with W62 and W63, as illustrated by the occupation rate of the Xo4 molecules at the protein surface during the simulations (Table 3-4).

Concerning the three extensions, the alcohol moiety of the TbXo4-OH variant presents more flexibility to explore the proteins surface and to create weak links, favouring versatile intermolecular interactions, in particular with D119 and Q121 from the symmetric protein as well as with N103 through weaker hydrogen bonds. On the other side the sulfonate and ammonium moieties seem to explore less the surface of the protein.

Also, arginine 5 (R5) shows that it may interact with one of the picolinate moiety in TbXo4, TbXo4-OH, TbXo4-SO_3^- and TbXo4-NMet2 with occupation rates of 0.95, 1.00, 0.77 and 0.99, respectively. Also, N103 can be involved in hydrogen bonds with the carboxylate of a picolinate moiety in TbXo4-SO_3^- and TbXo4-$NMet_2$.

Table 3-4 *Occupation rate of TbXo4 variants in HEWL during the molecular simulation.*

	TbXo4	TbXo4-OH	TbXo4-SO_3^-	TbXo4-$NMet_2$
W62	0.74	1	0.99	1
W63	0.99	1	1	0.95
L75	1	0.99	0.98	0.97
D101	1	1	1	1
N103	0.67	1	1	0.92
A107	-	0.64	0.98	0.74
D119	0.81	1	0.13	0.21
Q121	-	1	-	0.09
R125	-	0.10	0.16	0.25
R5	0.95	1	0.77	0.99
W123	0.83	0.18	-	0.29

In addition, the binding energy of the variants at the interface of the HEWL were evaluated by MMPBSA analysis, an approach that permits to estimate protein-ligand binding affinity by calculating the associated binding free energy. A positive binding energy corresponds to a repulsive interaction whereas a negative binding energy refers to an attractive interaction. The result of this analysis shows that all the variants possess a total positive binding energy, with a preference for TbXo4 (19.4 kcal·mol^{-1}) and TbXo4-OH (20.3 kcal·mol^{-1}) as they possess lowest binding energy hence less repulsive interactions, followed by TbXo4-$NMet_2$ (42.3 kcal·mol^{-1}) and then TbXo4-SO_3^- (48.3 kcal·mol^{-1}) and with standard deviation, for each binding energy, between 5 and 10 Kcal·mol^{-1}. Those results seem to be in agreement with the experimental results and highlight additional interaction sites, such as N103, A107 and R5 for TbXo4. Those residues were not directly observed in the crystal structure as potential residues involved in the interaction. The crystal structure represents the final state of the crystallization process, so it cannot be excluded that these residues are also involved during the formation of the crystal packing mediated by TbXo4. Moreover, the crystal was obtained in a certain crystallization condition. Thus, those interaction sites may be observed in structures resulting from crystals obtained from other crystallization conditions. Indeed, it would be interesting to compare these HEWL

results with data from crystals obtained in other crystallization conditions in presence of TbXo4.

3.4 Conclusion

The analysis conducted on thirteen Xo4 variants show that, TbXo4-OH variant with a propanol arm on the free nitrogen of the triazacyclononane macrocycle, presents the best results by providing a high rate of crystallization hits as well as additional hits when compared to the reference molecule TbXo4. In addition to a direct interaction with acidic residues and similarly to TbXo4, this variant is able to interact with different types of amino-acids, by generating hydrophobic interactions (CH-π) via its macrocycle (TACN) moiety, interact with positively charged amino acids such as arginine by one of its carboxylate function presented on the picolinate moiety. Also, it seems to bind better at the surface of the protein, due to its steric hindrance, compared to other variants. It was also confirmed by molecular dynamic simulations that TbXo4-OH, due to its alcohol group (donor/acceptor role), creates weak interactions while exploring the proteins surface, as its presents more flexibility comparing to the other variants.

For these purpose, we propose this variant as second generation of crystallophore.

However, these preliminary results will have to be consolidated by an evaluation of each of the variants on a larger panel of proteins. It would also be interesting to exploit unique crystallization conditions provided by each variant to even better describe the influence of each moieties on the binding of these molecules. Finally, it would be interesting to end a study I have initiated during my PhD aimed at evaluating the nucleation properties of TbXo4 using a set of 8 commercial proteins chosen to cover different molecular weight and assemblies as well as different pI values.

CHAPTER 4

CHARACTERISATION AND UNDERSTANDING OF $TbXO_4$ INTERACTIONS

4. CHARACTERISATION AND UNDERSTANDING OF TBXO4 INTERACTIONS

In this chapter, we will decipher TbXo4 binding sites at the protein surface in order to determine the predominant role of certain amino acids in the interaction. We also wanted to confirm that these amino acids are also crucial to the nucleating effect induced by TbXo4.

To do so, we have used a mutagenesis approach; an approach that consist of site-directed mutagenesis (single point mutation) for crystal engineering. The choice of this approach is based on a recent study that has shown that it is possible to carry out mutations in the conserved domain of a nanobody in order to allow fluorochromes or heavy metals to settle there (Hansen et al., 2017).

In addition, another approach (SERp) is to use the surface entropy reduction (SER) technique. The SER protocol, available as a web server, consists in identifying residues present in the wild-type protein sequence that may present high surface entropy and may hamper the crystallization of the considered protein. Then, it proposes mutations to remove these high surface entropy residues in order to improve the crystallizability of the protein (Goldschmidt et al., 2007). This approach, combined with the mutagenesis approach, was used by S. Engilberge when he studied TbXo4 binding sites on the surface of Protease 1 (PhP1), one of protein models that possessed high number of unique crystallization conditions (Engilberge et al., 2017). In his thesis, Sylvain showed that the single point mutations of key residues reduced the number of unique hits. He also evaluated that some of these residues are those proposed by the SERp server, indicating that one effect of TbXo4 binding could be to reduce the entropy of negatively charged residues.
The idea was thus to confirm these preliminary observations. For that, we have used two other model proteins. However, the SERp server is no longer supported and no longer requires authentication. As a consequence I focus on the use of the mutagenesis approach in this work.

The rationale is to determine if one (or few) amino-acid motif favourable to TbXo4 binding exist. This approach may allow, in the future, the creation of TbXo4 binding site ensuring the interaction of the complex with a protein of interest, thus favouring the crystallization of the latter.

Therefore, in the present chapter, using the mutagenesis approach, we will present the results obtained on two protein models:

1) Adenylate Kinase from *Methanothermococcus thermolithotrophicus* (MthAdkA) that forms a trimeric assembly and presents an original binding mode and asymmetry of binding with TbXo4 (Engilberge et al., 2019; Roux et al., 2021). Also, it presents lots of negatively-charged amino acids (Asp and Glu) on its surface (Chapter 2, Table 2-1).

2) Lysozyme from Hen Egg (HEWL) a commercial protein provided by Roche (reference: 10837059001), that possesses a large number of unique crystallization conditions induced by TbXo4 (Engilberge et al., 2017; chapter 3 Figure 3-2, 3-3 & 3-8).

Our approach was to use HTX-Lab platform, in order to track the crystallization rate of each mutant of our proteins of interest (in the absence and with TbXo4) and compare the results with the one observed with the wild type protein. A decrease in the number of hits, in particular of the unique hits of TbXo4, would confirm the importance of the targeted amino acid in the crystallization process. Thereafter, we have confirmed the absence or the presence of TbXo4 binding sites for each model proteins and their mutants respectively, by exploiting the anomalous contribution of TbXo4. Finally, molecular dynamic simulations were performed by our collaborators at ENS-Lyon (Elise Dumont and Natacha Gillet) to analyse our two model proteins and their mutants, in order to figure out the different TbXo4 binding modes and the energies governing the interactions between the complex and the surface of the considered protein.

4.1 Model 1: MthAdkA

4.1.1 MthAdkA native structure analysis and design of the mutants

The first model protein is Adenylate kinase (AdkA, ATP/AMP phosphotransferase, EC 2.7.3.4, ~20 KDa). The protein was initially purified natively from the anaerobic

marine archaea *Methanothermococcus thermolithotrophicus* (native MthAdkA) by Tristan Wagner (Max Planck Institute, Bremen), a collaborator in this project since 2016 (Engilberge et al., 2019; Vögeli et al., 2018).

As described in the introduction and summarized in Figure 4-1 a, Tristan Wagner's native approach consists of purifying proteins of interest from cultures of the marine microorganism *Methanothermococcus thermolithotrophicus*, which is their native host, in order to decipher their role in microbial metabolic pathways. He uses crystallization as an ultimate purification step and regularly applies TbXo4 to crystallize these proteins from enriched fractions that contain not only one protein (Figure 4-1 b). He also exploits TbXo4 for phasing (Engilberge et al., 2019; Roux et al., 2021).

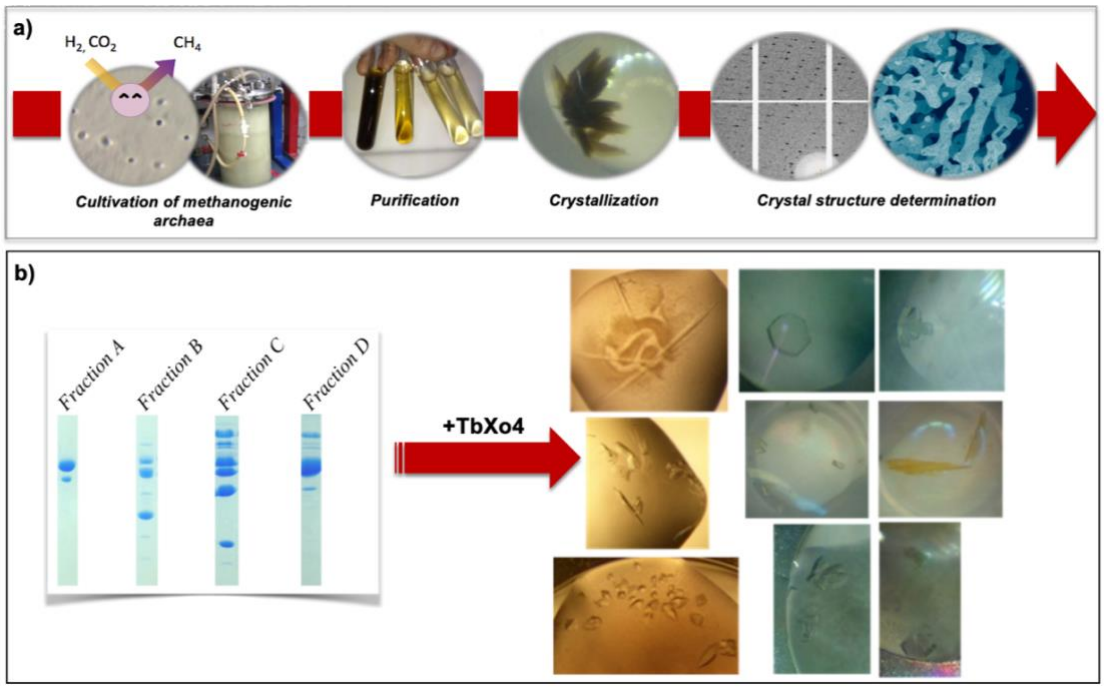


Figure 4-1 *a) Native approach by Tristan Wagner where proteins are isolated from their native host to decipher metabolic secrets (Vögeli et al., 2018). b) SDS page gel of the fractions obtained by native purifications from Methanothermococcus thermolithotrophicus and obtained crystals of proteins from these fractions in presence of 10 mM TbXo4. Native MthAdkA was present in Fraction B (Engilberge et al., 2019).*

The structure of native MthAdkA was obtained from diffraction data of a crystal of the protein co-crystallized in the presence of 10 mM TbXo4 with a crystallization condition consisting in 25% PEP629, 100 mM HEPES pH= 7.5, 50 mM $MgCl_2$ and soaked in a cryo-protectant solution containing 50 mM TbXo4 (Engilberge et al., 2019), leading to

perform an *ab initio* phasing with sequence identification from the experimental map. A trimeric assembly was identified in the asymmetric unit corresponding to the biological assembly. An original TbXo4 binding mode was observed (Roux et al., 2021)(Figure 4-2):

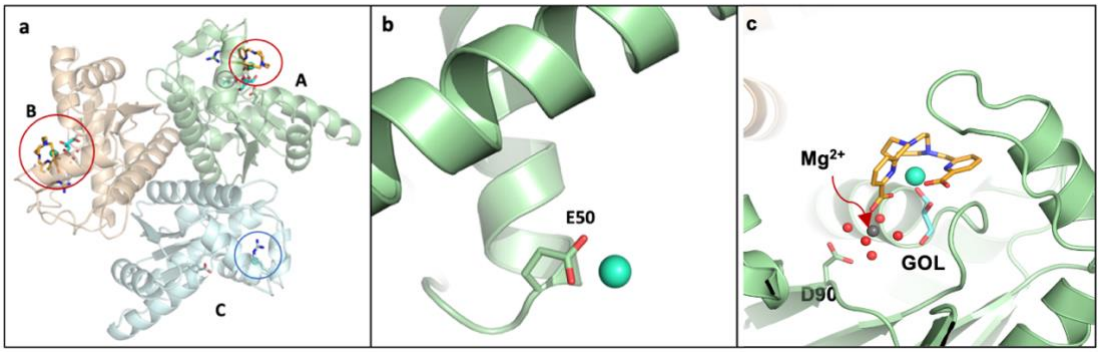


Figure 4-2 *Structure of AdkA co-crystallized with 10 mM TbXo4 and soaked in 50 mM solution of TbXo4 (PDB ID 6HF7) a) View of the MthAdkA trimer showing TbXo4 binding asymmetry. b-c) Interaction sites of TbXo4.*

i. Two Tb binding sites with low occupancies were determined: one was found in monomer A only (close to glutamate E50) (Figure 4-2 a grey circle & b) and a second low binding site was found only in monomer C near E29. For both sites, we hypothesized that they may be traces of a labile interaction resulting from the crystallization process induced by TbXo4. However it is not possible to exclude that they may also result from the soaking step in 50 mM TbXo4.

ii. Two strong Tb binding sites show an interaction with Aspartate D90 mediated by a hydrated Mg2+ ion. It also involves a molecule of glycerol (Figure 4-2 c). It is interesting to remind that TbXo4 is a cationic complex and, while observing a Tb^{3+} direct coordination by D90, we observed a hydrated cation involvement. This was already described in the case of the protein FprA with the presence of a calcium ion (Engilberge et al., 2018). These two strong binding sites were determined only in two monomers of the trimer (Figure 4-2 a red circle & c). The analysis of the same binding site on the third monomer shows that it is more occluded by the crystal packing. As a consequence, we have hypothesized that these strong binding sites can be the result of the soaking step rather than of the co-crystallization process.

The observations made rise the importance of collecting diffraction data from a MthAdkA crystal co-crystallized with 10 mM TbXo4 without any soaking. For that, I recombinantly expressed and purified MthAdkA wildtype protein (MthAdkA WT), to be able to get sufficient material for attempting to understand the mechanism of crystallization of MthAdkA induced by TbXo4. The expression and purification protocol of the protein are reported in Chapter 2.

After purification, I checked that MthAdkA WT still crystallized by reproducing the conventional crystallization condition found with the natively purified protein (PEP629, 100 mM Tris pH= 8.5, 50 mM $MgCl_2$). The crystallization diagram of MthAdkA WT shows that the recombinant protein behaves similarly to the native protein (Figure 4-3).

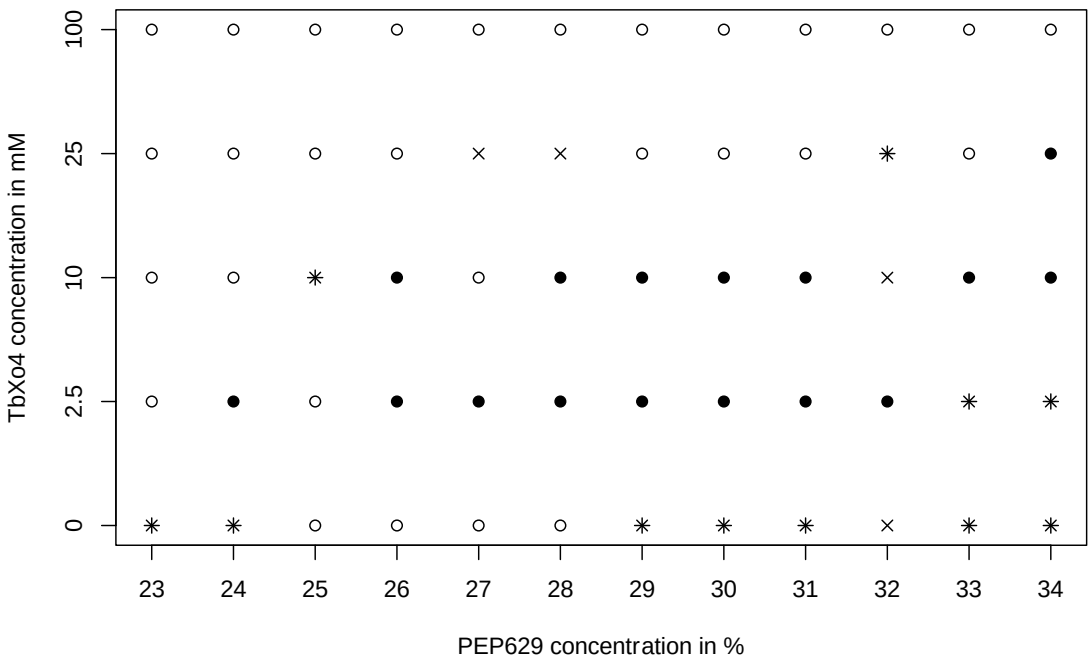


Figure 4-3 *Crystallization diagram of 10mg/ml of MthAdkA WT in the conventional crystallization condition used (Engilberge et al., 2019): PEP629, 100 mM Tris pH= 8.5, 50 mM $MgCl_2$. ° represents condition with clear drop crystals, • represents condition with single crystal, * represents microcrystals or/and crystalline precipitates and ˣ represents precipitation.*

After confirmation that the recombinant wildtype protein MthAdkA WT behave similarly in term of crystallization with respect to the natively produced protein, and in an attempt to decipher the mechanism of MthAdkA crystallization by TbXo4, I have expressed and purified two series of MthAdkA mutants (Table 4-1). The following paragraph first

summarizes the mutant choice and the detailed results for each of them will be presented afterward.

1. Serie 1: This series relies on the observations described previously and based on the detailed analysis of TbXo4 binding in the native AdkA structure. Consequently, we have chosen to replace the residues E50 (Mutant 1 or M1) and D90 (Mutant 2 or M2), related to the Tb binding sites, by an Alanine residue.

2. Serie 2: The second series of mutations was mainly based on molecular dynamics simulations of AdkA performed by our collaborators at ENS-Lyon (E. Dumont & N. Gillet). In brief, binding of TbXo4 to the trimeric AdkA was captured using molecular dynamics, with an extended sampling of 10 replicates of 200ns. The results from the MD simulations reveal 6 hotspots that are likely to interact with the TbXo4 based on the time spanned by the complex interacting with a particular residue divided by the total simulation time (per-residue normalized contact probability ρ). The residues presenting high contact time are: E29, E45, E49, E54, E136 and E185, respectively (Figure 4-4).

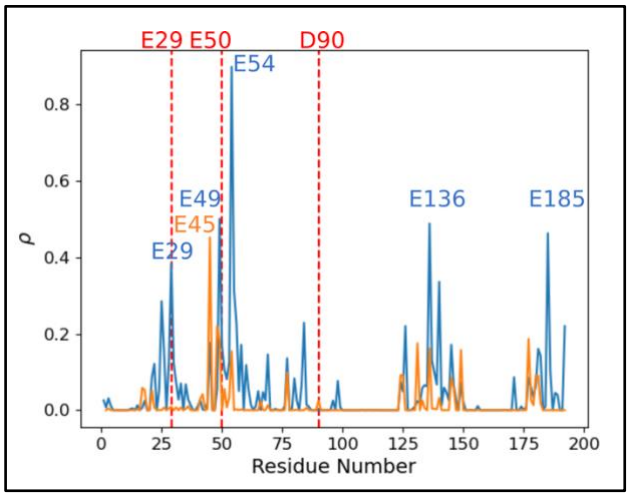


Figure 4-4 *Per-residue normalized contact probability ρ of TbXo4 (i.e. the time spanned by the ligand interacting within residue divided by the total simulation time) for MthAdkA, starting from the position from crystallographic structure (orange, 2 TbXo4 in the simulation box) or random position (blue, 10 TbXo4 in the simulation box). The crystallographic sites E29, E50 and D90 are identified with vertical red coloured dashed lines.*

It is worth noting that D90 was not listed while the minor site in native AdkA structure E29 is and that the second minor site E50 can be considered as a trace of the interaction with either E49 or E54, as these 3 residues are spatially close.

From the simulation results, we chose first to focus on residues 54 and 136 as they present the two highest ρ values. We also decided to modify our mutation strategy to limit potential mutant production issues and consequently we have substituted glutamate for glutamine and in the case of residue 54, we also evaluated the potential role of the side chain length by mutating it to aspartate. In summary, mutations were E54 to Q (Mutant 3 or M3), E54 to D (Mutant 4 or M4) and E136 to Q (Mutant 5 or M5).

Finally, based on preliminary results obtained on M1 to M5, we decided to design a hexa-mutant including the following mutations: E54Q, E49Q, E136Q, E185Q, E29Q and E149Q (Mutant 6 or M6). The choice to mutate E149 to Q was made as anomalous Fourier peaks close to E149 were observed in several MthAdkA structures (Table 4-2).

All the mutants (summarized in Table 4-1) were expressed and purified using a protocol similar to the one used as described in Chapter 2.

Table 4-1 *List of MthAdkA mutants and the corresponding mutations*

	MthAdkA mutants	Mutations
Serie 1	M1	E50A
	M2	D90A
Serie 2	M3	E54Q
	M4	E54D
	M5	E136Q
	M6	E29Q, E49Q, E54Q, E136Q, E149Q, E185Q

4.1.2 Crystallization output comparisons between WT/mutant MthAdkA.

All the mutants as well as MthAdkA WT (as a reference) were evaluated at the HTX platform (596 conditions) and the crystallization results of the mutants were compared with the MthAdkA WT one.

After purification, all samples were concentrated at 10 mg/mL with the addition of 10% glycerol, in order to be in the same conditions as the native protein. Then, these samples were used to dissolve the TbXo4 powder to a final concentration of 10 mM. Following the preparation of the samples, automated crystallization assays were performed in the absence and in the presence of TbXo4, similarly to the protocol used to highlight crystallophore nucleating properties (Engilberge et al., 2017).

Automated crystallization assays were complemented by the determination of crystallization diagrams for each mutant. As shown for MthAdkA WT, I exploited the PEP-based crystallization condition used for the structure determination of the native MthAdkA. It should be noted that this crystallization condition is not fully a TbXo4 unique one as microcrystalline precipitates are observed in the absence of TbXo4 (Figure 4-3).

Figure 4-5 summarizes the results of HTX crystallizations with Serie 1. M1 shows a small decrease in the number of native crystallization hits (unique native + common hits) compared to the MthAdkA WT protein, with no significant variation in the total number of TbXo4 hits, including unique hits. Mutant M2 shows no significant variation in the total number of crystallization hits with and without TbXo4. Thus, only mutation M1 (E50A) have an effect on the crystallization of the native protein. However, this mutation and the M2 (D90A) one do not hamper the action of TbXo4 as it induces 53 TbXo4 unique hits, which is significantly similar to MthAdkA WT (49 TbXo4 unique hits) .

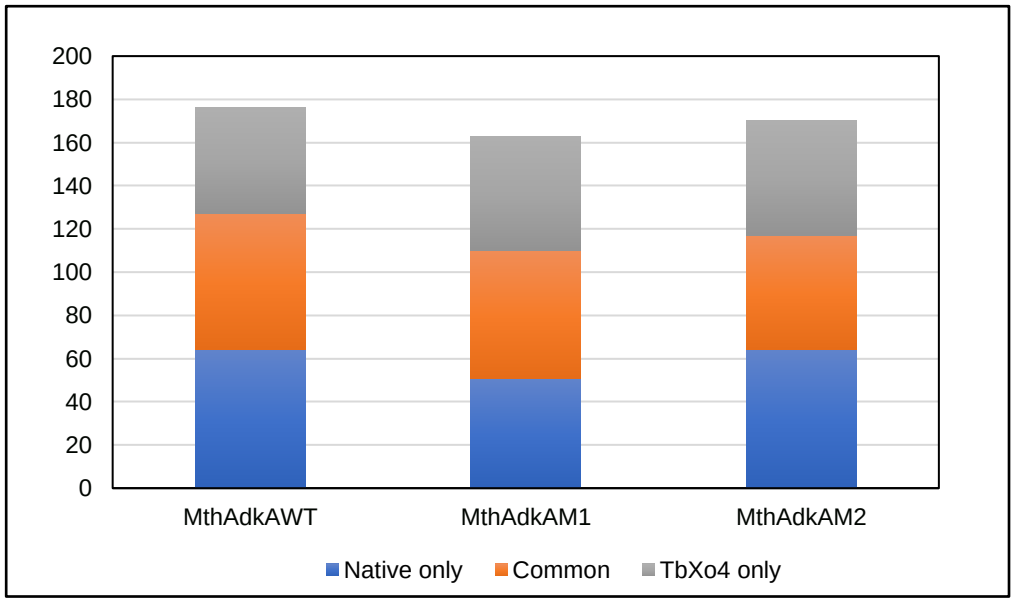
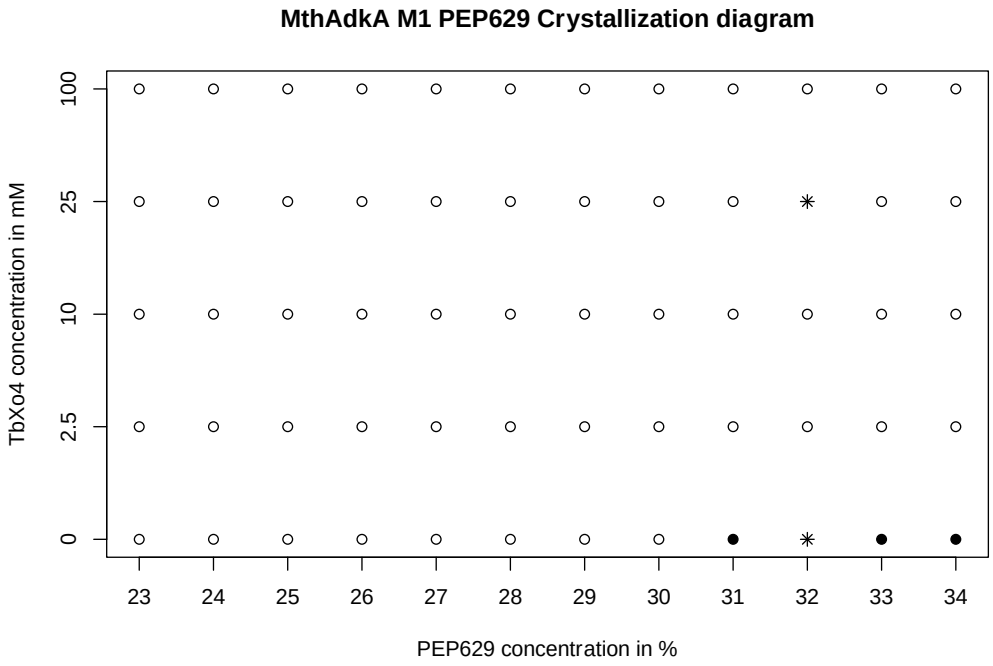


Figure 4-5 *Results of the automated crystallization screening of MthAdkA WT and mutants from Serie 1 (M1 & M2) in the absence/presence of 10 mM TbXo4 (observation after 30 days): unique native hits in blue, common hits (native and TbXo4) in orange, unique TbXo4 hits in grey.*

Following, as depicted in Figure 4-6, we determined the crystallization diagrams for the two mutants in the presence of PEP629 as precipitant and the results were compared with the one of MthAdkA WT shown in Figure 4-3. In this particular crystallization condition, it shows that M1 does not crystallize well in presence of TbXo4 (Figure 4-6, a) as most of the drops with TbXo4 remained clear. The native mutant M1 (without TbXo4) crystallize at high concentration of the precipitant (PEP629). As for M2, only the addition of 2.5 mM TbXo4 produces crystals at precipitant concentrations higher than those required for crystallisation of the native mutant protein (0 mM TbXo4) and it produces precipitation from the addition of 25 mM TbXo4. Unlike the HTX crystallization assay results, this shows that the two amino acids E50 and D90 are important for the interactions with TbXo4 and for inducing the nucleation, at least in the PEP-based crystallization condition.

a)



b)

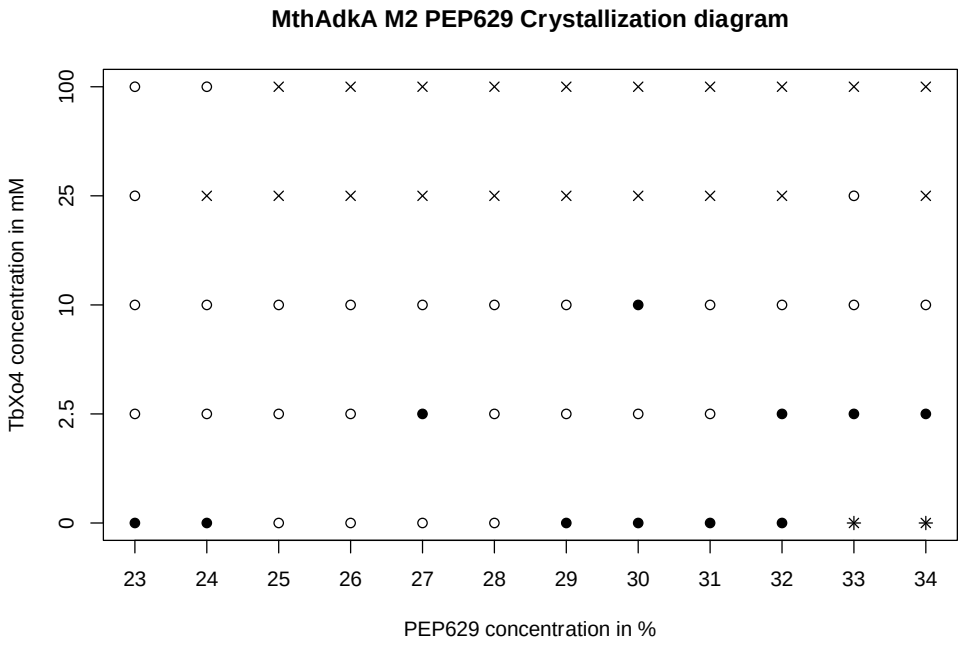


Figure 4-6 *Crystallization diagrams of 10mg/ml of MthAdkA (a) M1 (E50A) and (b) M2 (D90A) in the conventional crystallization condition (Engilberge et al., 2019): PEP629, 100 mM Tris pH= 8.5, 50 mM $MgCl_2$. ° represents condition with clear drop crystals, • represents condition with single crystal, * represents microcrystals or/and crystalline precipitates and × represents precipitation.*

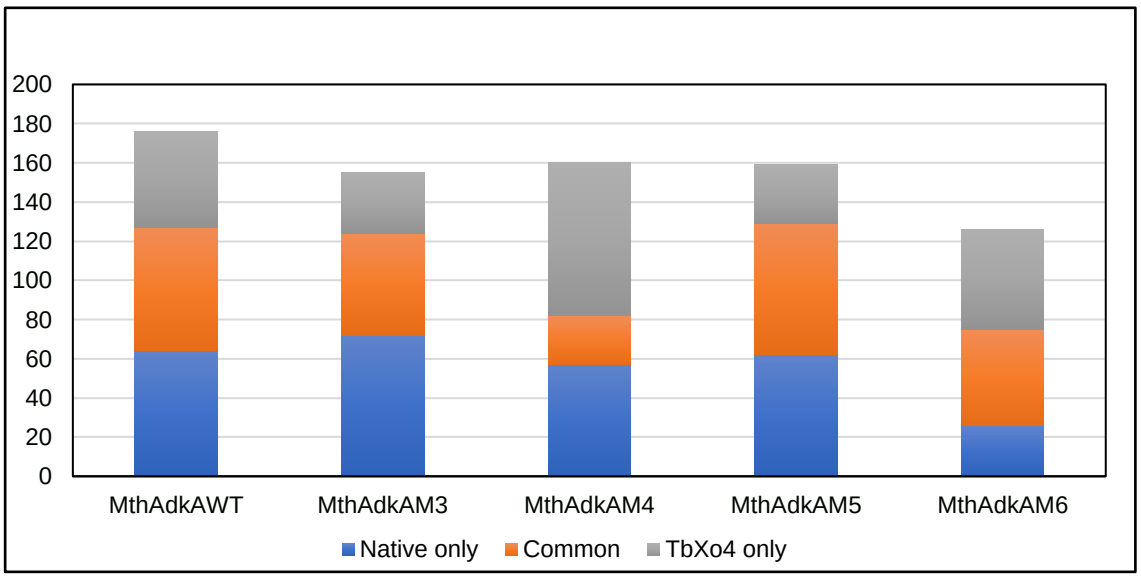


Figure 4-7 *Results of the automated crystallization screening of MthAdkA WT and its mutants from Serie 2 (M3, M4, M5 & M6) in the absence/presence of 10 mM TbXo4 (observation after 30 days): unique native hits in blue, common hits (native and TbXo4) in orange, TbXo4 unique TbXo4 hits in grey.*

Concerning the second mutant Serie, Figure 4-7 shows the results of automated crystallization assays. Mutant M3 shows a small decrease in the number of crystallization hits of the protein with TbXo4, compared to the MthAdkA WT protein, with no significant variation in the total number of hits without TbXo4. As for M4, a decrease in protein crystallizability in the absence of TbXo4 is observed. However, no significant variation in the total number of crystallization hits with TbXo4, with a significative increase in the number of TbXo4 unique hits (Figure 4-7). Thus, the substitution of E54 to an uncharged amino acid (especially Q) have an effect on the crystallizability of the protein in presence of TbXo4 with no effect on the crystallizability of the native protein. In general, the substitution of glutamic acid to aspartic acid (D) leads to the destabilization of a protein as it decreases the unfolding transition temperature Tm (Lee et al., 2004). This may explain the decrease in the protein crystallizability observed for M4. However, as mentioned previously, this mutation did not affect the crystallizability of the protein in presence of TbXo4 due to the fact that TbXo4 interaction with the aspartic acid stabilize the protein, thus reducing the mutation effect on the crystallizability of the protein in presence of TbXo4.

Regarding M5, no significant variation of the number of crystallization hits of the protein with or without TbXo4 is detected. However, there is a small decrease in the number of TbXo4 unique conditions.

About M6, crystallization assay of this mutant shows a decrease in the total number of crystallization hits of the protein without TbXo4. However, no significant decrease in the total or unique crystallization hits of the protein with TbXo4 is observed, while we were expecting a significant decrease of this number thanks to the number of mutations introduced. Indeed, as mentioned before, the mutations studied in M6 have been suggested by the MD results as the residues that have the longest time of contact with TbXo4 and their modification should have induced major difference in term of crystallization. In addition, despite the presence of E54Q mutation in M6, that is also present in M3 and leads to a decrease of total and unique number of TbXo4 hits, the E54Q mutation does not lead to the same effect in M6 (Figure 4-8). This may be due to a mutation between the other five that mitigates the effect of E54Q mutation, allowing TbXo4 to interact with other interesting, not yet determined by crystallography, sites that induce protein crystallization, as MthAdkA presents lots of negatively charged amino-acids on its surface. Note that the MD simulation have proposed other binding sites of TbXo4 which present low ρ value (Per-residue normalized contact probability) but may be sufficient for MthAdkA crystallization induced by TbXo4.
Furthermore, there could also be a compensatory effect since for example E54 is spatially close to E50.

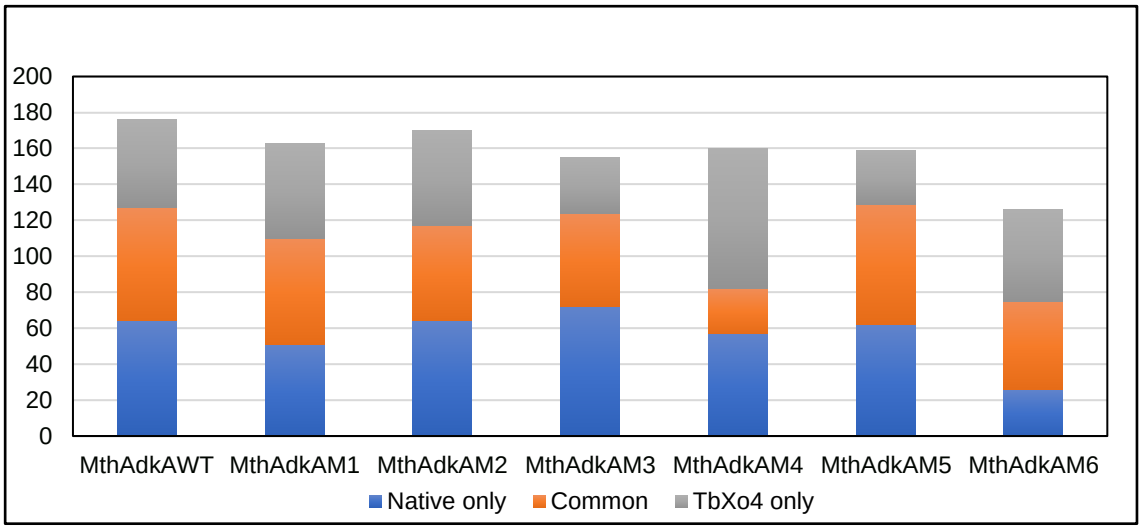


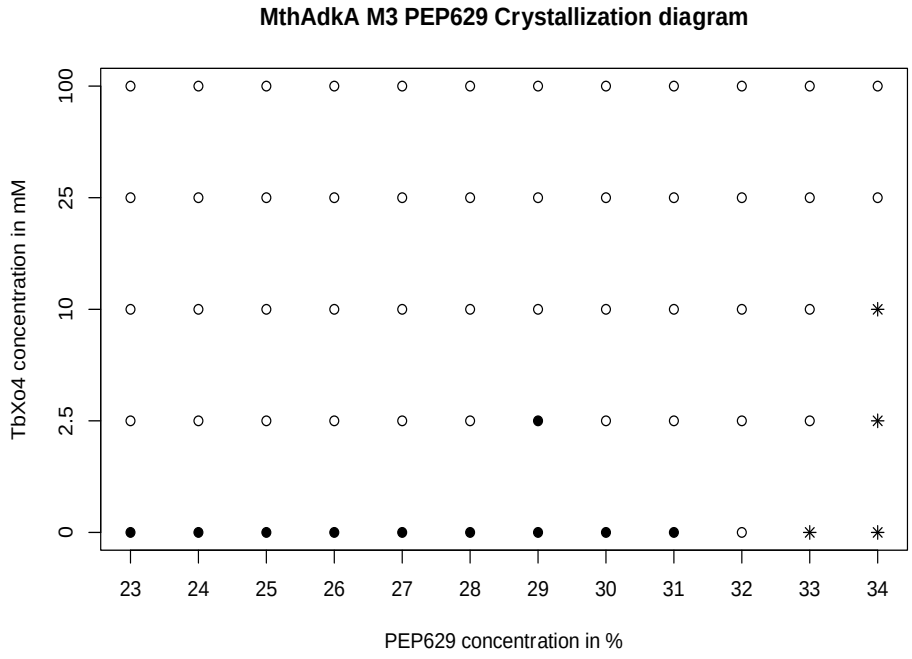
Figure 4-8 *Results of the automated crystallization screening of TbXo4 with MthAdkA WT and its mutants after 30 days: unique native hits in blue, common hits (native and TbXo4) in orange, TbXo4 unique hits in grey. In a red square marked the second mutation Serie 2. To facilitate analysis, results of Serie 1 are also indicated.*

Furthermore, like for the mutants in Serie 1, I have tested PEP-based crystallization condition (PEP629) on Serie 2 mutants (Figure 3-9). The results from the crystallization diagrams show that for M3, the mutant crystallizes in the absence of TbXo4 (0 mM TbXo4) similarly to MthAdkA WT. At the opposite, the addition of TbXo4 almost does not produce crystals (Figure 4-9 a), in that particular crystallization condition (PEP629). For both M4 and M5 mutants, the addition of TbXo4 allows production of crystals up to 10 mM of TbXo4 for both mutants (Figure 4-9 b-c) similarly to what observed for MthAdkA WT (Figure 4-3), with crystalline precipitates obtained for the crystallization of the native mutant protein unlike MthAdkA WT.

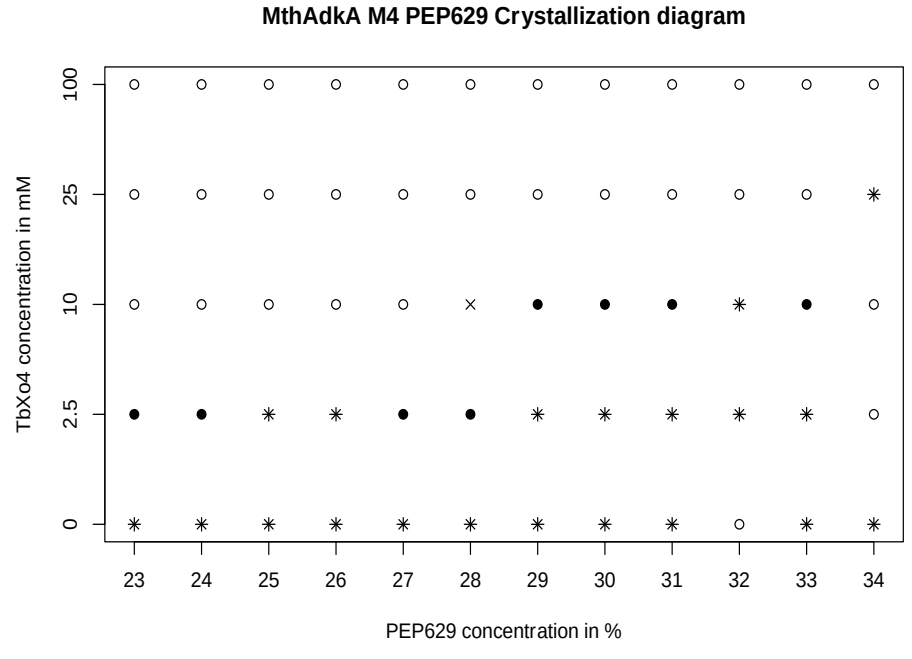
Finally for M6, no crystals were observed whatever TbXo4 is present or not.

From PEP629 crystallization diagrams on Serie 2 mutants, the results show that E54 is important for the interactions with TbXo4 and for inducing the nucleation in this particular crystallization condition and that the mutation of E54 to aspartic acids (D), obviously, does not affect much the crystallizability of the protein in presence of TbXo4 comparing to the mutation to glutamine (Q). This also confirms the interplay between the role of specific amino acid and the influence of the crystallization components on the TbXo4 binding as already shown (Engilberge et al., 2018; Jiang et al., 2020; Roux et al., 2021).

a)


MthAdkA M3 PEP629 Crystallization diagram
TbXo4 concentration in mM
0
2.5
10
25
100
23
24
25
26
27
28
29
30
31
32
33
34
PEP629 concentration in %

b)


MthAdkA M4 PEP629 Crystallization diagram
TbXo4 concentration in mM
0
2.5
10
25
100
23
24
25
26
27
28
29
30
31
32
33
34
PEP629 concentration in %

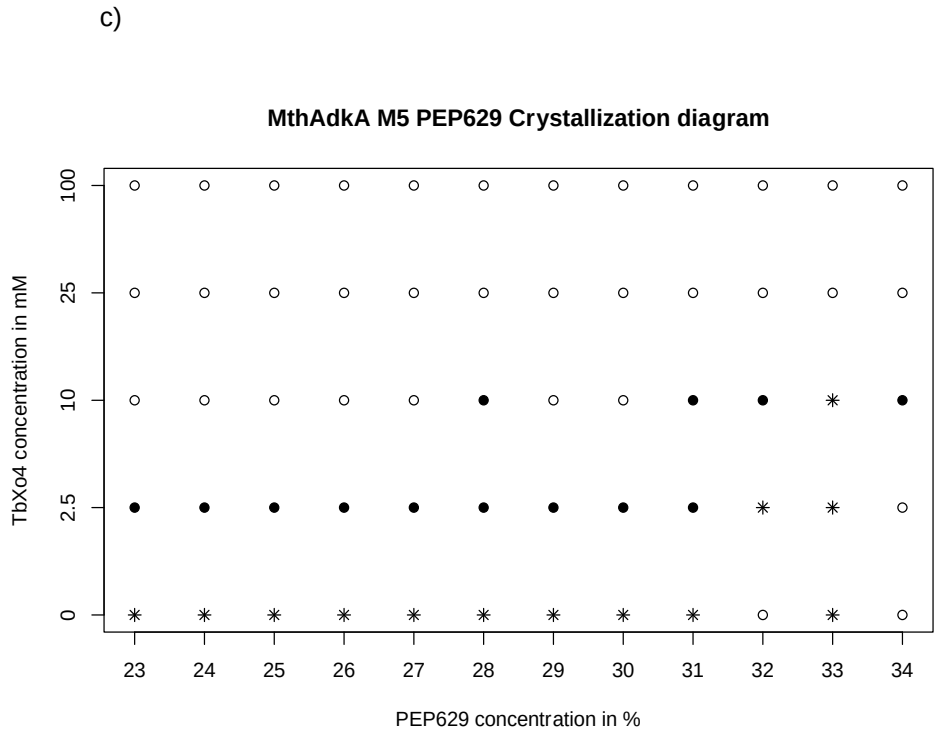


Figure 4-9 *Crystallization diagrams of MthAdkA (a) M3 (E54Q), (b) M4 (E54D) and (c) M5 (E136Q) in the conventional crystallization condition used in previous work: PEP629, 100 mM Tris pH= 8.5, 50 mM $MgCl_2$. ◦ represents condition with clear drop crystals, • represents condition with single crystal and * represents microcrystals or/and crystallin precipitates.*

4.1.3 WT and mutant MthAdkA structure analysis

We exploited several crystals for both WT and mutants MthAdkA and used the anomalous contribution of TbXo4 to compute anomalous Fourier synthesis rather than performing complete refinement of the structures. As mentioned previously, we wanted first to verify if the binding sites observed in native AdkA (PDB 6HF7) were not the result of the soaking step at 50 mM TbXo4. Second, we have evaluated the influence of the different mutations on TbXo4 binding.

Before describing the obtained results, I should remind here a behaviour of Fourier synthesis. Figure 4-10 illustrates two situations corresponding to different binding situations of two anomalous scatterers and the resulting influence on site detection. In Situation 1, the presence of a strong binding site (high site occupancy) results in a higher RMSD of the anomalous Fourier map compared to Situation 2. As a consequence, the minor site (Peak2) will not be detected in the first situation while it

will be in Situation2. This explains why it can be beneficial to exploit crystals from different crystallization conditions and/or with different space groups to analyse the binding of molecule such as TbXo4.

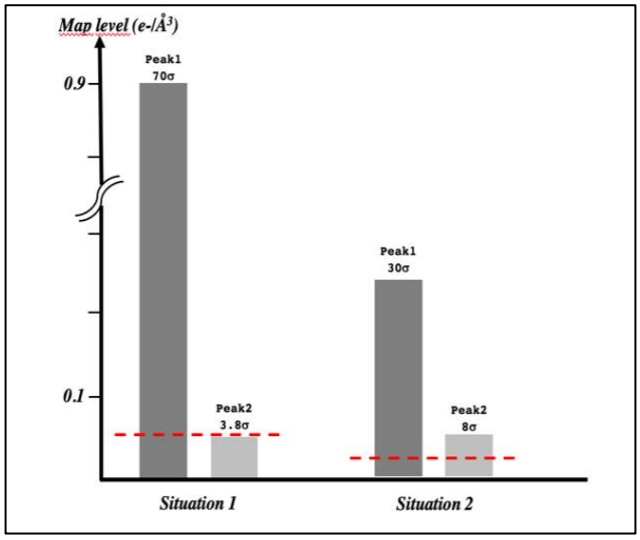


Figure 4-10 *Illustration of the influence of a strong binding site on the anomalous Fourier signal to detect minor sites. The red dotted line indicates the 4σ threshold used for peak detection.*

Also, the diffraction data were all collected at the Tb L_{III} absorption edge to maximize the anomalous contribution. A molecular replacement run was performed using 6HF7 as starting model to precisely place the model. This ensures optimal determination of the phases for Fourier synthesis. The anomalous Fourier synthesis is computed using an in-house script using CCP4 programs but can also be obtained with software like ANODE (Thorn & Sheldrick, 2011).

The results of the anomalous Fourier analysis in MthAdkA WT and mutants structures, as well as the corresponding crystallization conditions, data resolutions and space groups are reported in Table 4-2.

Therefore, we have first collected diffraction data on MthAdkA WT co-crystallized with 10 mM TbXo4 and without any soaking, in the conventional crystallization condition PEP629 (Table 4-2; WT-1). The obtained structure possess identical space group (P 4 2_1 2) to the native structure obtained from native purification (PDB: 6HF7). The calculated anomalous Fourier synthesis shows no peak in the previously described

interaction zone involving D90 (Engilberge et al., 2018) neither near E29. We also detected new sites (Table 4-2, WT-1) that correspond to some proposed by MD simulations (E54, E136, E148 & E185).

On the basis of the PEP629 crystallization condition, we can conclude that the major binding site (D90) observed in the structure of MthAdkA natively produced arises rather from the soaking than from the co-crystallization step (Figure 4-2).

However, we also exploited MthAdkA WT crystals obtained in two other crystallization conditions. From the second structure of MthAdkA WT obtained at high concentration of MPD and at pH=10.5 (Table 4-2; WT-2), we noticed, in the first instance, the presence of anomalous Fourier signal positioned at the binding site D90 which we have concluded earlier that it results from the soaking in the PEP629 condition. Thus, from this result, we realize that, not only the soaking, but also the crystallization condition (concentration of the precipitants and pH) may have an effect on the observation of binding sites for TbXo4. We also detected new sites that were obtained by MD simulations (E54, E136 & E185). For the third MthAdkA WT structure (WT-3) (Table 4-2), the calculated anomalous Fourier synthesis shows similar results to the one obtained with the diffraction data with PEP629, co-crystallized 10 mM TbXo4 and without soaking, with new observed binding sites (E49, E54, E136, E149, E185 & E30), in addition to E50 and E29.

Therefore, from all the MthAdkA WT diffraction data obtained in different crystallization conditions combined, we determined that TbXo4 binding site at D90 in PEP629 condition results from soaking but that this binding site also depends on the crystallization condition. Additionally, we detected all the binding sites obtained by Molecular dynamic simulation.

Furthermore, we have collected diffraction data on some mutants is order to verify if the mutations may have possibly eliminated TbXo4 binding sites (Table 4-2).

The obtained structure of MthAdkA mutants also possess identical space group (P 4 2_1 2) to the native structure obtained from native purification (PDB: 6HF7).

For M1, by comparing it with WT-3 which were obtained in the same condition, we observe the disappearance of the binding site near E50 in monomer A due to the substitution of this amino acid.

As for M3, the diffraction data were obtained in PEP629 condition. The anomalous Fourier synthesis shows the disappearance of the binding site near E54 in monomer

B, as well as the absence of signal near E185 for monomer A and C. This shows the importance of the E54 amino acid for the interaction with TbXo4.

Regarding M4, the diffraction data were obtained in another condition than the WTs. However, comparing the anomalous Fourier synthesis signal obtained in M4 with all WTs ones, it seems that the substitution of E54 to an aspartic acid has allowed the appearance of binding sites at this position in all monomers, in addition to the other binding sites found which are similar to the ones observed in WTs' data. Therefore, the mutation E54 to D shows the importance of this amino acid for the interaction with TbXo4, as concluded above, but also that the nature of the amino acid has an influence. Indeed, this may explain the increase of TbXo4 crystallization unique hits observed (Figure 4-8).

For M5, the computed anomalous Fourier synthesis from M5 diffraction data shows that the mutation of E136Q does not affect the binding sites observed previously in WT. However, it appears the presence of anomalous Fourier signal near D90 in monomer A and B even though the crystallization condition of M5 is PEP629 without soaking but with a higher concentration of PEP629 compared to the WT ones.

Finally, for M2 and M6, the obtained crystals were tested and diffracted either at very low resolution that does not allow us to obtain the anomalous Fourier signal or have high mosaïcity.

Table 4-2 *Crystallization conditions of MthAdkA WT and mutants and anomalous Fourier signal positions (*signal between 4 and 6 σ).*

Protein	Crystallization condition	Resolution (Å)	Space Group	D90	E49	E50	E54	E136	E149	E185	D181	E29	E30
Native	10 mM TbXo4, soaking 50 mM condition PEP629	1.96	P 4 2_1 2	A/B		A						C	
WT-1	10 mM TbXo4 17%, PEP629 100 mM Tris pH=8.5, 50 mM MgCl2	2.3	P 4 2_1 2			A/C	B	A/C*	A/B/C	A/B/C			
WT-2	10 mM TbXo4, MPD 43%, CAPS pH= 10.5, 10 mM MgCl2	1.99	P 4 2_1 2	A/B		A/C	B	A		A/B*	A/B*/C		C*
WT-3	10 mMTbXo4, PEG3350 26%, BisTris propane pH=8.5, 200 mM sodium fluoride 10 mM MgCl2	2.3	P 4 2_1 2		A/C*	A/C*	B*	A	A/B/C	A*/B/C*		C	B*C*
M1	10 mMTbXo4, PEG3350 34%, BisTris propane pH=8.5, 200 mM sodium fluoride 10 mM MgCl2	1.99	P 4 2_1 2		C*		A*/B*	A	A/B/C	A/B/C		C*	A*/C*
M3	10 mMTbXo4, PEP629 28%, 100 mM Tris pH=8.5, 50 mM MgCl2	2.09	P 4 2_1 2			A		A	A/B/C	B*			
M4	10 mM TbXo4, PEG400 32%, Tris pH=8.5	2.2	P 4 2_1 2			A/C	A/B/C	A	A/B/C	A/B/C	A*/B*/C*	C*	C*
M5	10 mM TbXo4, 29%PEP629, 100 mM Tris pH=8.5, 50 mM MgCl2	2.3	P 4 2_1 2	A/B		A/C	B	A	A/B/C				

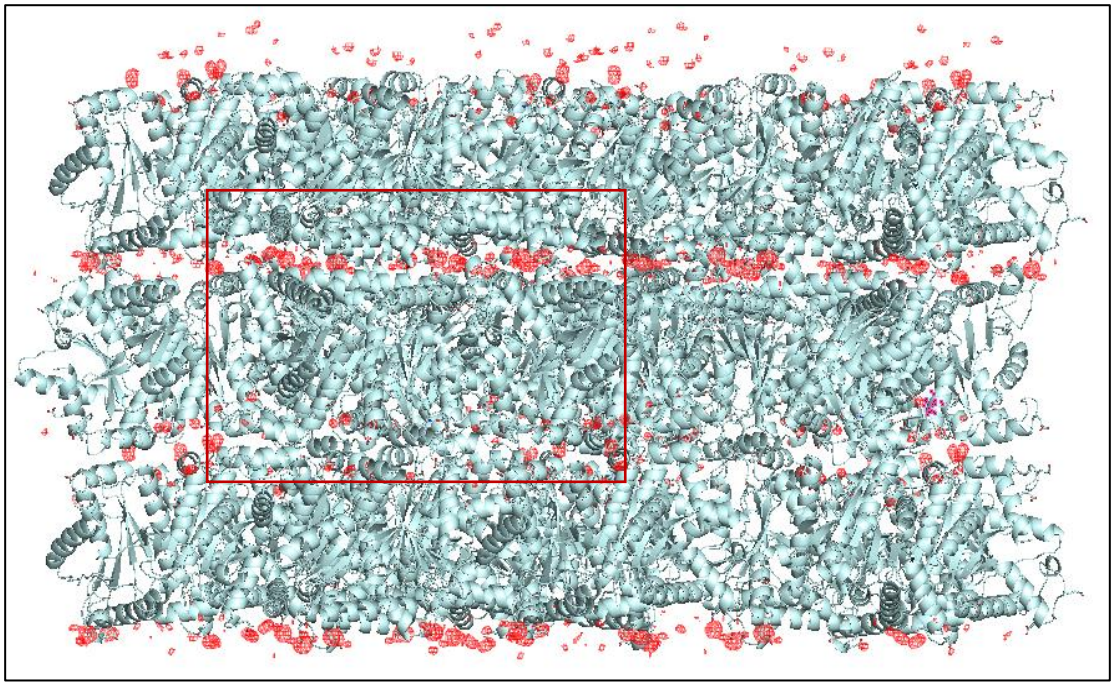
4.1.4 MthAdkA conclusion

From the conventional crystallization diagram and the crystallization assay, we may say that M1 and M2 are interesting mutants for MthAdkA. Both mutants M1 and M2 has shown that they are important for the interactions with TbXo4 and for inducing the nucleation in the PEP-based crystallization condition. However, from the HTX crystallization assay, these two mutants did not affect TbXo4 actions on the protein as they did not affect TbXo4 total and unique hits. Therefore, M1 and M2 do not offer much information to understand TbXo4 interaction in MthAdkA.
More importantly, mutants M3 and M4 have shown to have a significant effect in MthAdkA. These two mutants have shown their significant impact on the crystallizability of the native MthAdkA, in case of M4, as well as the influence of the mutation of E54, depending on the type of the mutation, on the crystallization of MthAdkA in presence of TbXo4. In addition, from the collected data, it has also shown the role of specific amino acids such as E49, E136 and E185, in addition to E54, and the influence of the crystallization conditions on the TbXo4 binding.

Moreover, by generating the crystal packing of MthAdkA WT-1, it shows that the gap in the packing area is completely filled with TbXo4 (Figure 4-11 a), with the presence also of some of the amino acids that are considered to interact with TbXo4 on the surface of MthAdkA WT (Figure 4-11 b), according to the observations from the structures (Table 4-2) and MD simulation results, indicating the role of different amino acids in the interaction with TbXo4. Thus, the constant presence of several major and minor sites that interact with TbXo4.

As a result, this protein model is not the best model to study TbXo4 interactions by site-directed mutagenesis.

a)



b)

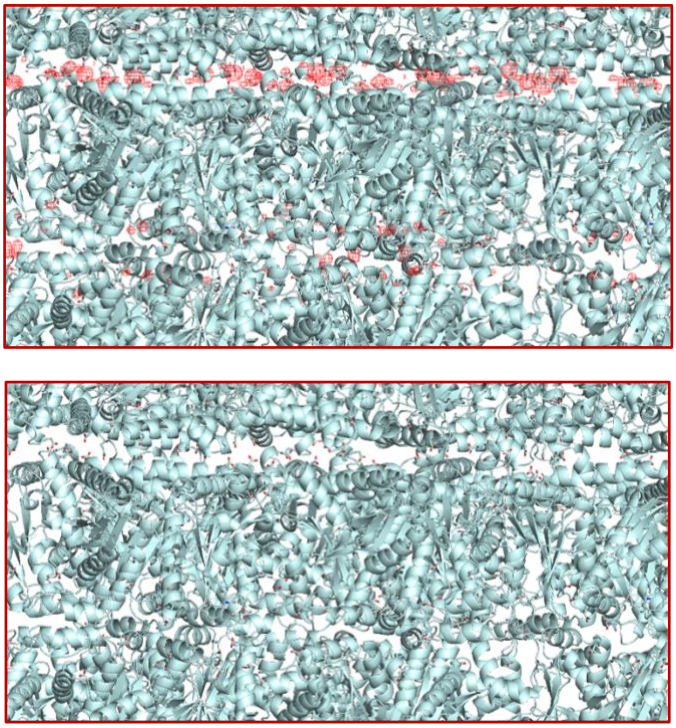


Figure 4-11 *Crystal packing of MthAdkA WT-1 (Table 4-2) and a) the anomalous Fourier map in red representing Tb atom positions b) Residues that interacts with TbXo4 on the surface of the protein; on top the anomalous Fourier map with the residues that interacts with TbXo4 and on the bottom the residues.*

4.2 Model 2: Egg-White Lysozyme (From the eggs to the structures: Egg hunt)

The second model used extensively for the characterisation of crystallophore interactions is Hen Egg-White Lysozyme (HEWL). HEWL is a commercial protein that possesses a large number of unique crystallization conditions induced by TbXo4 during HTX-Lab screenings, which was also used for evaluating the nucleating properties of crystallophore variants.

As seen before in Chapter 3, the structure of HEWL with 10 mM TbXo4 indicates that residue D101 is the main interacting residue through a direct coordination of its carboxy oxygens with the terbium ion of the crystallophore. Hence, it rises the interest in studying the effect of D101 mutation on the nucleation of the lysozyme in presence of TbXo4.

Unfortunately, due to lysozyme antimicrobial function which consists in the hydrolysis of the peptidoglycan components of bacterial cell walls, the expression of lysozymes in *E. coli* is not feasible as it leads to rapid cell lysis during its expression, therefore leading to poor yield of the protein (Fischer et al., 1993).

Luckily, we found that the mutation D101G naturally occurs in Turkey Egg-White Lysozyme (TEWL). Accordingly, we have searched for other lysozymes that may have suitable mutations in the cluster that binds TbXo4 (i.e. the vicinity of D101 or other residues involved in secondary interactions) and for which egg availability can be managed. We found Quail Egg-White Lysozyme (QEWL) which naturally has both G102V and N103H mutations and Peacock Egg-White Lysozyme (PEWL) which presents G102R mutation (Figure 4-12). In addition, we also considered two other lysozymes from Emu and Ostrich Egg-White (EEW and OEW) that possesses G102K mutation. However, it turns out that these two species do not express the desired lysozyme isoform which is lysozyme C (C for Chicken or conventional type). OEW expresses the G isoform of lysozyme (goose-type), which has no significant similarity to the C isoform (Maehashi et al., 2012; Weaver et al., 1984), while EEW has a very low level of expression of lysozyme G (Maehashi et al., 2012). Hence, we were not able to experimentally exploit the G102K mutation.

As shown on Figure 4-12, in addition to the direct mutation of D101 provided by TEWL, the selected egg-white lysozymes allow to evaluate the influence on TbXo4 binding

and nucleating properties of (i) bulkier residues in the vicinity of D101 (QEWL) and (ii) a positively charged amino acid next to D101 (PEWL). The mutations presented in each EWL species are presented in Table 4-3.

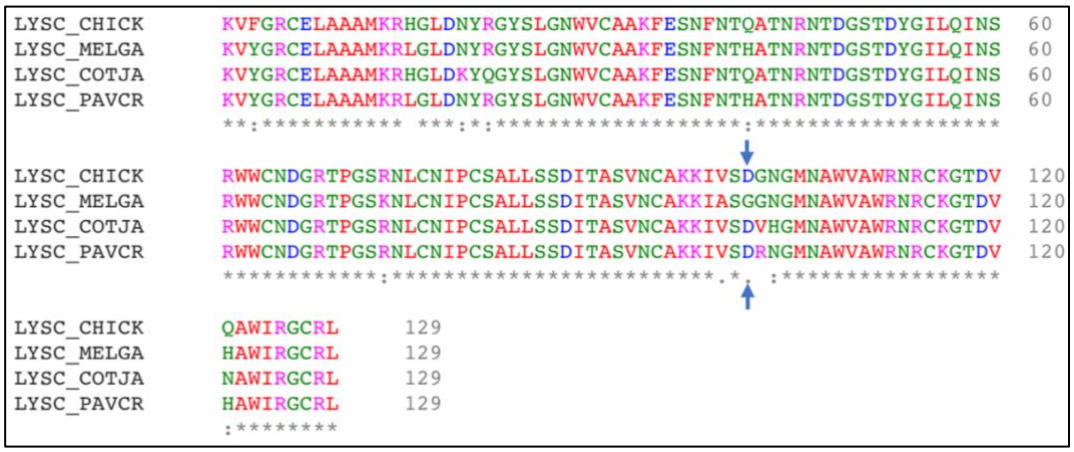


Figure 4-12 *Multiple sequence alignment by Clustal Omega of Lysozymes from Hen Egg-White (LYSC_CHICK), Turkey Egg-White (LYSC_MELGA), Quail Egg-White (LYSC_COTJA) and Peacock Egg-White (LYSC_PAVCR).*

Table 4-3 *EWL naturel mutations compared to HEWL*

Egg-White Lysozymes (EWL)	Mutations (HEWL as reference)
TEWL	F3Y,H15L,Q41H,R73K,V99A,D101G,Q121H
QEWL	F3Y,N19K,R21Q,G102V,N103H,Q121N
PEWL	F3Y,H15L,Q41H,G102R,Q121H,

Therefore, the three selected proteins TEWL, QEWL and PEWL were extracted and purified from eggs of *Meleagris gallapova* from local farms, *Coturnix japonica* from local stores and *Pavo cristatus* species from Chateau de Vizille Park, respectively, according to the protocol described in Chapter 2.3.

In addition, we used a similar molecular dynamics simulation (MD) approach to AdkA, using EWL structures from the PDB (part 4.2.3), for blind evaluation, as they do not present TbXo4 in their structure, in order to analyse the interaction sites between the protein and the crystallophore. Only PEWL structure that we determined in presence of TbXo4 were used in this analysis, as an experimental structure of PEWL was not available, yet. In Addition to EWL used for crystallization, we benefit from the PDB in

order to test other lysozyme such as the Ostrich Egg-White Lysozyme (OEWL) and Softshell Turtle Egg-White Lysozyme (STEWL). The MD experiments were performed by our collaborators at ENS-Lyon (E. Dumont & N. Gillet).

4.2.1 Effect of the natural mutations on the crystallization of EWL in the presence of TbXo4

After purification of EWL (Chapter 2.3), protein samples were concentrated at 20 mg/mL in a storage solution consisting in 50 mM NaCl, 50 mM Tris-HCl pH=7.5. The protein solution were used to dissolve TbXo4 powder to a final concentration of 10 mM, as previously described for HTXLab experiments. Noted that for this experiment, HEWL, purchased from Roche, was prepared in the same buffer as the one of the other purified lysozymes.

Automated crystallization assays of the four samples were performed at the HTXLab platform (Grenoble), in order to evaluate the effect of the mutations on the crystallization rate of the different lysozymes in presence of 10 mM TbXo4. The results of these assays are shown in Figure 4-13.

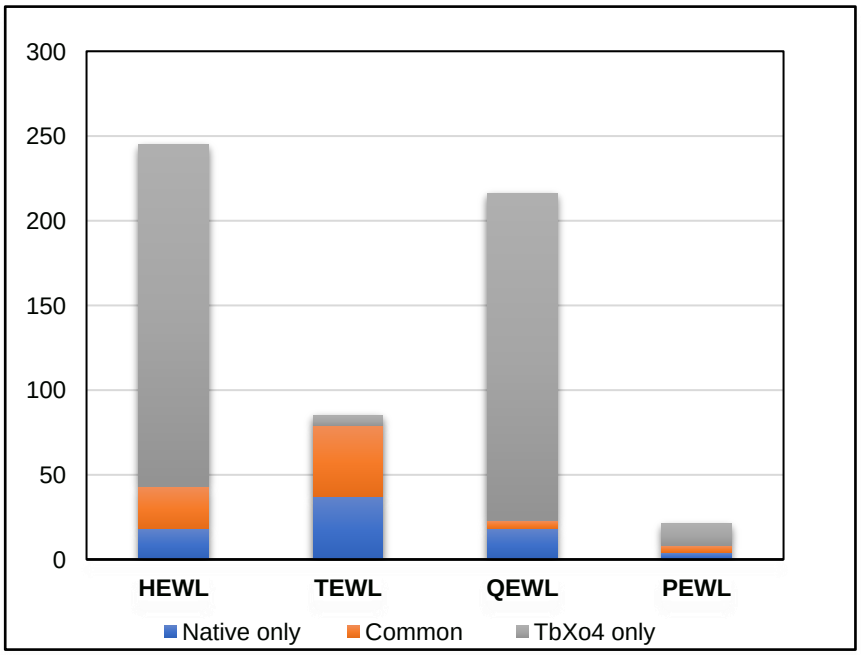


Figure 4-13 *Results of the automated crystallization screening of EWL with TbXo4 after 30 days: number of hits for native protein in blue, common hits (native and TbXo4) in orange, TbXo4 unique hits in grey.*

Similarly to what was observed previously (Chapter 3; Figure 3-3 a), the results obtained from the crystallization assays show that TbXo4 presents the highest nucleating effect on HEWL as it covers 58% of the native hits and induces 202 additional unique hits, with a minor influence of the change of protein buffer (Buffer B') (Chapter 2.3) on the number of crystallization hits in absence and in presence of TbXo4 comparing to the values obtained with the protein prepared in H_2O (Native-avg and TbXo4-avg in Chapter 3 Figure 3-3). At the opposite, while covering 53% of the native hits, TbXo4 induces only 6 unique hits in the case of TEWL. This clearly demonstrates the preponderant role of D101 in the nucleating effect of TbXo4 in EWL, as TbXo4 stabilise this residue, hence favouring the crystallization of HEWL. It is also worth noting that D101 partly hampers the crystallizability of HEWL due to the fact that D101 presents high surface entropy. Indeed, TEWL presents the highest native hit number due to surface entropy reduction by D101 natural mutation.

For QEWL, we observed a reduction in the number of native crystallization hits, showing that the naturally mutated amino acids in QEWL (K19, Q121N and the two mutation near D101, G102V and N103H, with reference to HEWL) have an effect on the crystallization of the native protein. However, these mutations, especially the ones near D101 (G102V and N103H), do not hamper the action of TbXo4 as it induces 193 unique hits, which is significantly similar to the number of hits observed for HEWL. In the present case, TbXo4 hits only cover 22% of the native hits. That may be directly related to the lower propensity of the native protein to crystallize rather than to a reduced nucleating effect of TbXo4 as shown with unique hits.

Concerning PEWL, we notice a major decrease in the number of native hits (-81%) compared to the number of native hits of HEWL as well as the huge decrease of the number of TbXo4 hits (-92.5%). This important decrease in the number of crystallization hits with and without TbXo4 was not expected, as PEWL presents five mutations in its sequence, four of which are the same natural mutations as in TEWL. As shown above these four mutations do not affect the crystallization of TEWL native form thus either in PEWL. Thus, the 5^{th} mutations in PEWL, G102R which is also near D101, is considered responsible for hindering the crystallisation of PEWL.

Therefore, from QEWL and PEWL crystallization assays, it shows that depending on the nature of the amino acid presented at the mutational site G102, an effect on the crystallization of the protein may be observed.

Furthermore, similarly to AdkA study, I evaluated the effect of the concentration of TbXo4 on the crystallization of EWL by realising crystallization diagrams. For that I exploited a conventional HEWL crystallization condition that consists of NaCl as precipitant with sodium acetate buffer at pH=4.6 (Figures 4-14 & 4-15). Note that these manual plates were made with 1+1+1 protocol mentioned in chapter 3.1.

The results from crystallization diagram of HEWL indicates (Figure 4-14 a), firstly, as previously described in Engilberge et al. (2017), TbXo4 strongly enlarges the crystallization diagram of HEWL. Also, it shows that the crystallization diagram of HEWL dissolved in EWL buffer is similar to the diagram of the protein dissolved in distilled water.

Regarding TEWL's results (Figure 4-14 b), in the condition exploited, TEWL crystallises much less compared to HEWL, in the absence and presence of TbXo4. Indeed, the addition of TbXo4 slightly enlarges the crystallisation zone of TEWL up to 10 mM, confirming a strong decrease in the nucleating efficiency of TbXo4 associated with the D101G mutation.

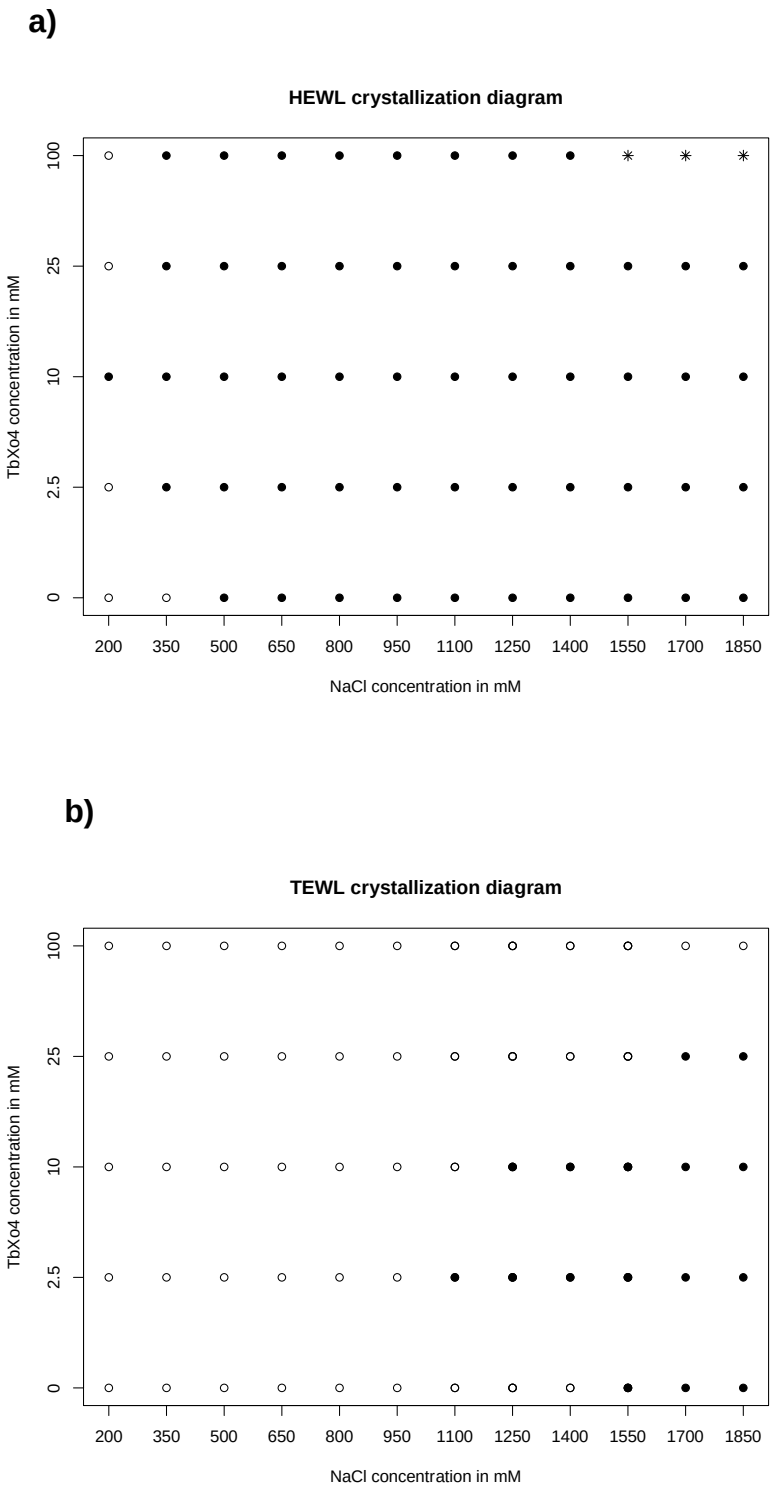


Figure 4-14 *Crystallization diagrams of (a) HEWL and (b) TEWL after 30 days with different TbXo4 concentrations (0, 2.5, 10, 25 and 100 mM). ◦ represents clear drops, • represents condition with single crystal, * represents microcrystals or/and crystalline precipitates and × represents precipitation.*

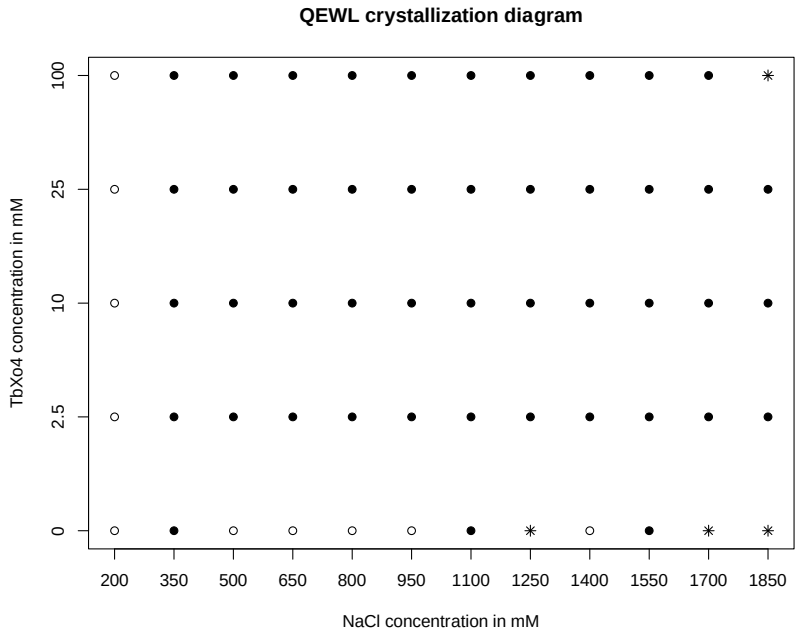
Unlike TEWL, QEWL shows a crystallization diagram similar to HEWL in the considered conventional condition (Figure 4-15 a), with a slight decrease in the crystallization of the native protein and without affecting the crystallization of the protein in presence of TbXo4. Hence, this confirms the results obtained from HTXLab assay on QEWL.

As for PEWL, a shift to the right of the crystallisation diagram is observed (Figure 4-15 b). The situation is even worse since the addition of TbXo4, regardless of its concentration, does not allow crystals to be obtained at precipitant concentrations lower than those required for crystallisation of the native protein (0 mM TbXo4).

Therefore, the results derived from the crystallization diagrams of the different EWLs confirm the conclusions drawn from the HTX experiments. Position D101 plays a dominant role in the nucleation effect of TbXo4 on EWL. Furthermore, mutation of position G102 to a positively charged amino acid has a clear deleterious effect on the crystallisation of the protein, while mutation to a bulkier residue has a limited effect.

In the following paragraph, I will evaluate the influence of the different mutations, on the binding mode of TbXo4 in the four EWL, by determining their crystal structures.

a)



b)

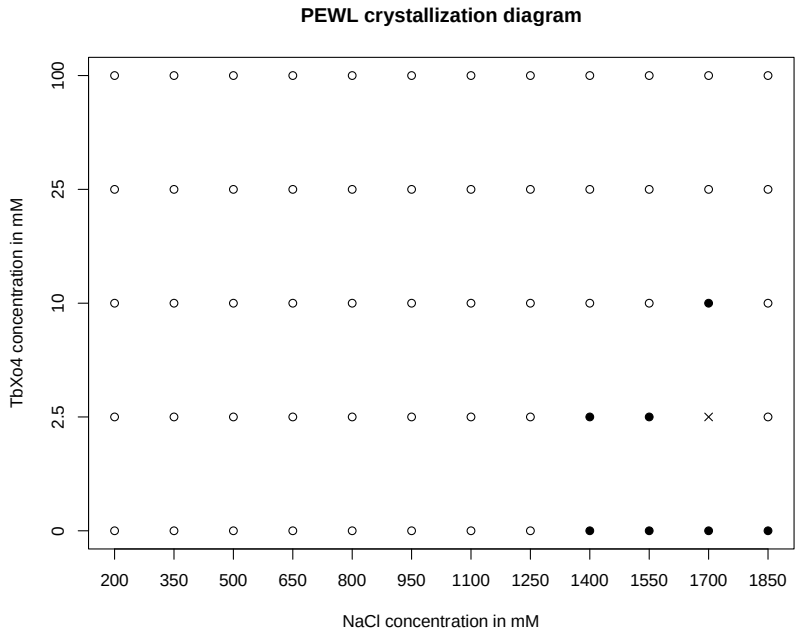


Figure 4-15 *Crystallization diagrams of (a) QEWL and (b) PEWL after 30 days with different TbXo4 concentrations (0, 2.5, 10, 25 and 100 mM). ◦ represents clear drops, • represents condition with single crystal, * represents microcrystals or/and crystalline precipitates and × represents precipitation.*

4.2.2 Evaluation of TbXo4 interaction with EWL

To evaluate the influence of the different natural mutations on the binding mode of TbXo4, we have determined the structures of the three natural mutants from crystals co-crystallized with 10 mM TbXo4 in the crystallization conditions reported in Table 4-4. Data collection were performed at L_{III} absorption edge of Terbium for all the crystals. Data quality and refinement statistics are given in Table 4-5 and 4-6. HEWL structure is the structure described in paragraph 3.3.1.

Table 4-4 Crystallization conditions of EWL.

Protein	Crystallization condition	Cryoprotectant
HEWL	800 mM Sodium Chloride, 100 mM Sodium Acetate pH= 4.6	1 M NaCl, 100 mM Sodium Acetate pH= 4.6, 30% PEG400
TEWL	1.85 M Sodium Chloride, 100 mM Sodium Acetate pH= 4.6	1.85 M Sodium Chloride, 100 mM Sodium Acetate pH= 4.6, 30% Glycerol
QEWL	800 mM Sodium Chloride, 100 mM Sodium Acetate pH= 4.6	1 M NaCl, 100 mM Sodium Acetate pH= 4.6, 30% PEG400
PEWL	1.9 M Sodium Formate, 100mM Bis-Tris pH=6	1.9 M Sodium Formate, 100 mM Bis-Tris pH=6, 30% Glycerol

From the first observation of the data quality and refinement statistics, QEWL and PEWL seem to present same group as HEWL ($P4_3\,2_1\,2$). Whereas, TEWL presents an hexagonal space group. Also, Q/PEWL present better diffracting behaviour than HEWL. This is shown by the integration statistics which indicate that QEWL and PEWL present better diffracting sample as illustrated by the I/sig(I) value at high resolution. This is also confirmed from the refinement of the structures that leads to lower overall B-factor, 23.68 and 16.58 $Å^2$ for QEWL and PEWL respectively, comparing to 32.33 $Å^2$ for HEWL. For TEWL, refinement is still to be completed.

Table 4-5 *Structural data collection and refinement statistics of HEWL and TEWL in the presence of 10 mM TbXo4.*

	HEWL 10 mM TbXo4	TEWL 10 mM TbXo4
Data collection		
Synchrotron-Beamline	SOLEIL-PX1	SOLEIL-PX2
Wavelength (Å)	1.64895	1.64895
Space group	P 4_3 2_1 2	P 6_1 2 2
Resolution (Å) [a]	37.49 - 1.70 (1.79 - 1.70)	41.45 - 2.15 (2.27 - 2.15)
Cell dimensions		
a, b, c (Å)	77.474, 77.474, 37.491	70.573, 70.573, 82.907
R_{sym} or R_{merge} (%) [a]	9.7 (33.6)	8.9 (55.6)
R_{pim}(%) [a]	2.6 (7.3)	1.6 (9.0)
$CC_{1/2}$ (%) [a]	99.8 (98.6)	99.8 (99.1)
I/σ(I) [a]	22.8 (9.0)	41.3 (16.6)
Completeness (%) [a]	99.7 (98.8)	97.7 (95.9)
Redundancy [a]	24.1 (23.5)	37.1 (38.2)
Number of unique reflections [a]	13059 (1848)	6866 (648)
Refinement		
Resolution (Å)	33.75 - 1.72	
Number of reflections	13015 (1848)	
R_{work} / R_{free} [b] (%)	16.99 / 20.96	
Number of atoms		
Protein	1019	
Ligands/ion	61	
Solvent	133	
Mean B-factors (Å^2)	32.33	
Protein	30.90	
Ligand/ion	50.30	
Water	38.19	
Ramachandran plot		
Favoured regions (%)	98.43	
Allowed regions (%)	1.57	
Outlier regions (%)	0.0	
R.m.s. deviations		
Bond lengths (Å)	0.017	
Bond angles (°)	1.49	

[a] Values in parentheses are for highest-resolution shell. [b] R_{free} was calculated as the R_{work} for 5% of the reflections that were not included in the refinement. [c] rmsd. root mean square deviation.

Table 4-6 *Structural data collection and refinement statistics of QEWL and PEWL in the presence of 10 mM TbXo4.*

	QEWL 10 mM TbXo4	PEWL 10 mM TbXo4
Data collection		
Synchrotron-Beamline	SOLEIL-PX1	SOLEIL-PX1
Wavelength (Å)	1.64895	1.64899
Space group	P 4_3 2_1 2	P 4_3 2_1 2
Resolution (Å) [a]	38.73 - 1.72 (1.81 - 1.72)	38.82 - 1.73 (1.82 - 1.73)
Cell dimensions		
a, b, c (Å)	77.460, 77.460, 38.160	77.642, 77.642, 37.708
R_{sym} or R_{merge} (%) [a]	5.5 (16)	5.1 (9.7)
R_{pim}(%) [a]	2.57 (5.07)	2.7 (4.5)
$CC_{1/2}$ (%) [a]	99.8 (99.3)	99.9 (99.8)
I/σ_I [a]	40.28 (14.2)	34.3 (15.8)
Completeness (%) [a]	99.98 (100)	99.9 (99.4)
Redundancy [a]	23.6 (22.6)	24.5 (23.9)
Number of unique reflections [a]	12824 (1237)	12500 (1208)
Refinement		
Resolution (Å)	38.73 - 1.72	34.72 - 1.73
Number of reflections	12830 (1237)	12500 (1207)
R_{work} / R_{free} [b] (%)	15.77 / 18.79	15.80 / 18.83
Number of atoms		
Protein	1018	1037
Ligands/ion	56	107
Solvent	159	147
Mean B-factors (Å^2)	23.68	16.58
Protein	22.23	14.97
Ligand/ion	26.63	15.13
Water	32.37	28.52
Ramachandran plot		
Favoured regions (%)	98.43	99.21
Allowed regions (%)	1.57	0.79
Outlier regions (%)	0.92	0.0
R.m.s. deviations		
Bond lengths (Å)	0.009	0.009
Bond angles (°)	1.08	1.05
Number of TLS groups	9	

[a] Values in parentheses are for highest-resolution shell. [b] Rfree was calculated as the Rwork for 5% of the reflections that were not included in the refinement. [c] rmsd. root mean square deviation.

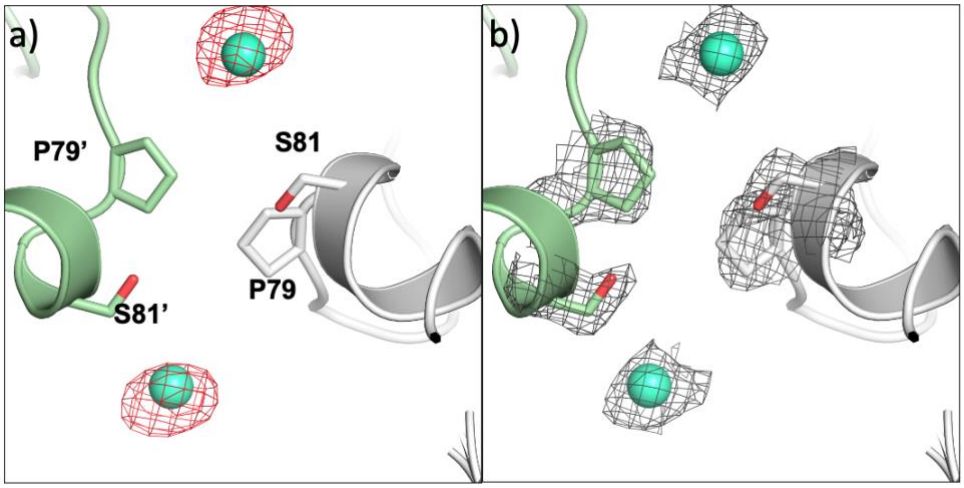


Figure 4-16 *Structure of TEWL co-crystallised with 10 mM TbXo4. a) anomalous Fourier map (contoured at 5 sigma) b) 2Fo-Fc electron density map superimposed to interaction sites of Tb (contour: 1 sigma). Tb atoms are represented as cyan sphere.*

By the same approach as MthAdkA, I checked the binding sites of TbXo4 in TEWL by the anomalous Fourier synthesis. The anomalous Fourier map from TEWL structure shows that the mutation D101G has eliminated TbXo4 interaction, as expected. No anomalous signal can be detected close to G101. However, a new site was discovered, in a common conserved domain within EWL near S81 (Figure 4-16), and which is not observed in H/Q/PEWL structures.

For QEWL, the comparison with the structure of HEWL with 10 mM TbXo4 from chapter 3.3.1 shows that TbXo4 seems to have mostly the same conformation in both structures (Figure 4-17 e,f) . However, the 2Fo-Fc electron-density map for the TbXo4 molecule in QEWL structure is better than in HEWL structure (Figure 4-17 b). This is associated to a higher occupancy of TbXo4 in QEWL than in HEWL, with occupancy values of 0.8 and 0.6 respectively, and to a less agitated terbium atom than the one in HEWL (average B-factor: 33.59 and 60.88 $Å^2$, respectively). Compared to TbXo4 binding site in HEWL structure, the binding site environment in QEWL is similar. The main binding site involves the aspartate D101 in QEWL, as well as residues such as W62, W63, L75, V102, H103 and D119, as confirmed by the corresponding 2Fo-Fc electron density maps (Figure 4-17 c). Similarly to HEWL, we observe a direct coordination of the terbium atom of the TbXo4 with the carboxylate moiety of aspartate 101 (Figure 4-17 d). The interaction distances between the oxygens and terbium are

2.6 Å which is slightly smaller than the distances observed in HEWL structure (2.7 Å), in addition to the weak interactions of V102 and H103 (CH-π ; 3.75 Å and NH-π ; 3.85 Å, respectively), may explain the better interaction of TbXo4 with QEWL surface.

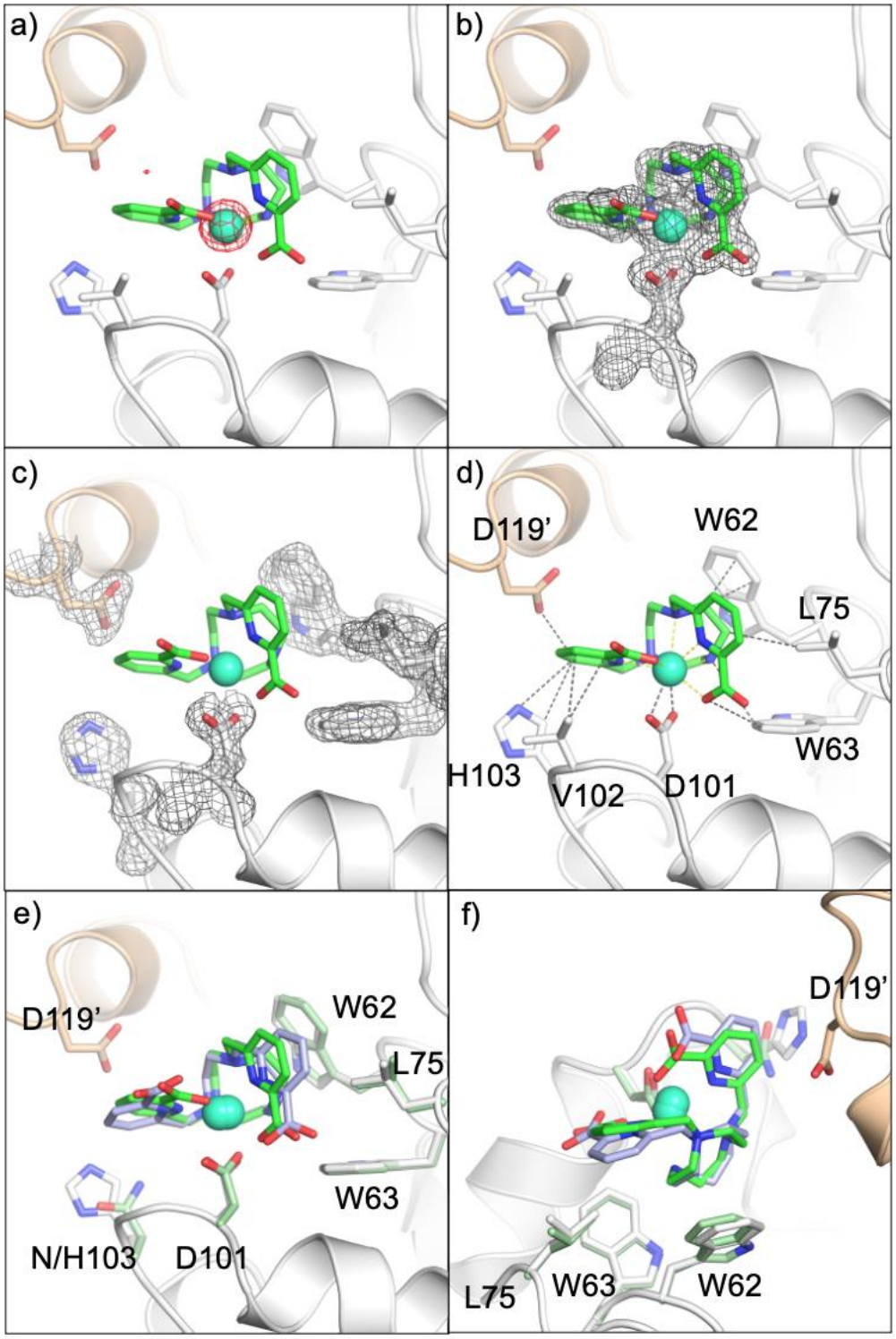


Figure 4-17 *Structure of QEWL co-crystallised with 10 mM TbXo4. a) anomalous Fourier map (contour: 5 sigma) b) 2Fo-Fc electron density map superimposed to TbXo4 molecule (contour: 1 sigma) c)) 2Fo-Fc electron density map corresponding to interaction sites of TbXo4 (1 sigma) d) Interaction sites of TbXo4. e-f) Superposition of TbXo4 in QEWL structure (TACN in green) and HEWL structure (palegreen) co-crystallized with 10 mM TbXo4 (TACN in lightblue). Tb atoms are represented as cyan sphere.*

Regarding the structure of PEWL, although the structure is related to a protein co-crystallized with 10 mM TbXo4, the binding site shows, unexpectedly, the presence of a dimer of TbXo4 between D101, W62 and W123 from an asymmetric molecule confirmed by the presence of two peaks in the anomalous Fourier map (Figure 4-18 a-b). Thus, the binding site is similar to the structure of HEWL co-crystallized with 100 mM TbXo4 (Engilberge et al., 2018). The 2Fo-Fc electron-density map coverage of TbXo4 in PEWL is good, even if it is not covering the whole second molecule (Figure 4-18 c-e). This is associated to lower occupancy of the two TbXo4 sites in PEWL (occupancy of site 1 & 2: 0.7/0.5) than in HEWL (0.85/0.7; PDB: 6F2I), with a less agitated terbium atom in HEWL (average B-factor of site 1 & 2 in PEWL: 23.07/25.92 $Å^2$ and in HEWL : 18.92/17.54 $Å^2$).

In comparison with the structure of HEWL with 100 mM TbXo4, the first TbXo4 seems to have only minor change in the conformation (Figure 4-18 e). Similarly to the dimer in HEWL, we observe a direct coordination of the terbium atom, of the first TbXo4 complex by the carboxylate function of aspartate 101 (Figure 4-18 d-f). The interaction distances between the oxygens and terbium are 2.38 and 2.52 Å respectively. In addition, W62 stabilizes the binding of TbXo4 in site 1 via a CH-π interaction between the macrocycle of the TbXo4 complex and the aromatic part of W62 (Figure 4-18 d).

The second TbXo4 complex is positioned at the interface between two protein chains (Figure 4-18 d) and on the contrary to the first molecule, it shows major changes in its orientation comparing to the one in HEWL structure (Figure 4-18 f). These changes in the conformations are due to the presence of G102R mutation that must have created steric hindrance, inducing a different conformation of the TbXo4 ligand than that obtained in HEWL. As a consequence, the second TbXo4 creates only an CH-π interaction with W123 which is 3.75 Å, and due to its change of orientation, it does not interact with K33 as in HEWL.

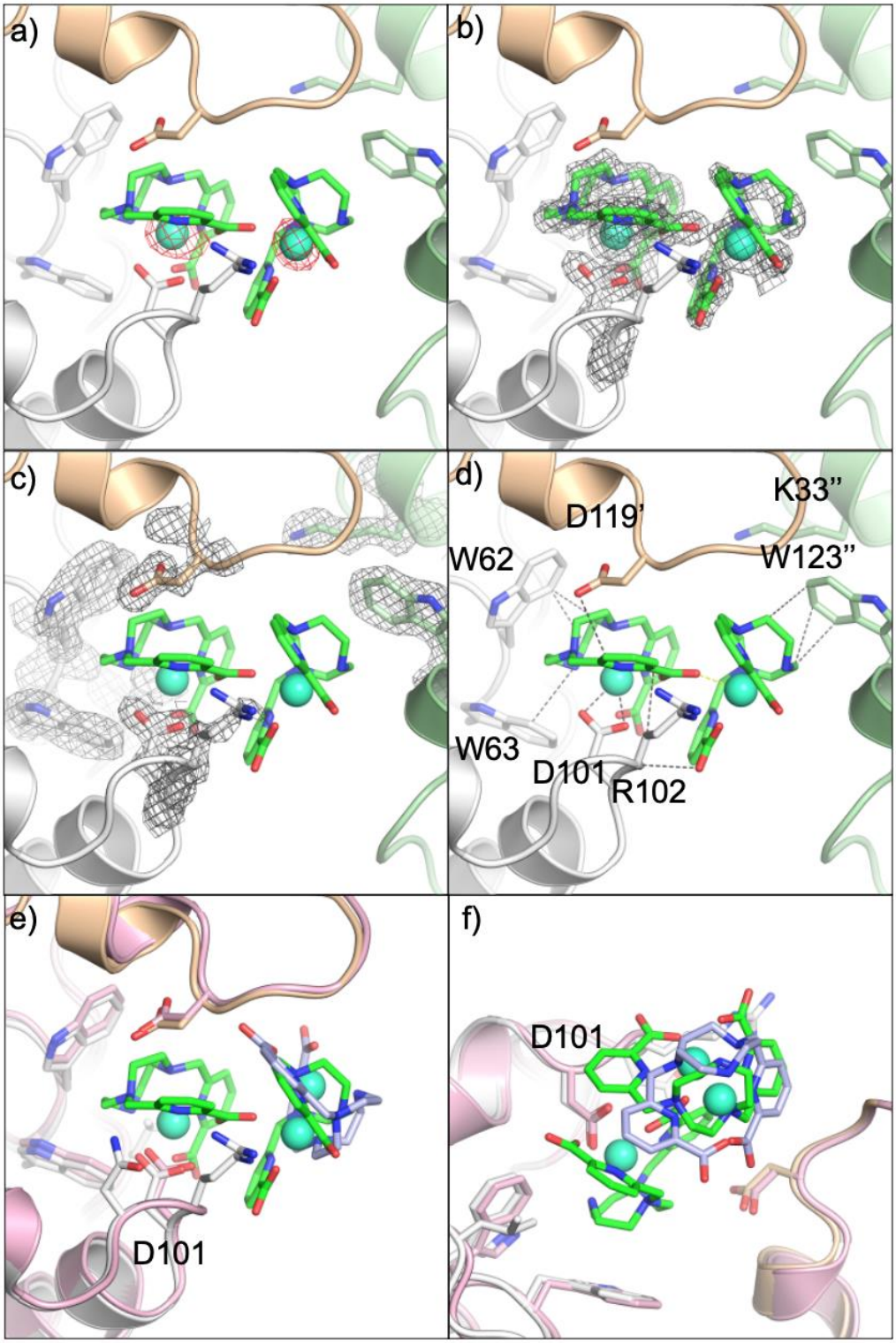


Figure 4-18 *Structure of PEWL co-crystallised with 10 mM TbXo4. a) anomalous Fourier map superimposed to Tb atoms (contour: 5 sigma) b) 2Fo-Fc electron density map superimposed to TbXo4 molecule (1 sigma) c)) 2Fo-Fc electron density map corresponding to interaction sites of TbXo4 (1 sigma) d) Interaction sites of TbXo4. e-f) Superposition of TbXo4 in PEWL structure (protein in grey, symmetry in wheat; TACN in green) and HEWL structure (pink, PDB ID 6F2I) in the presence of 100 mM TbXo4 site 2 (TACN in lightblue). Tb atoms are represented as cyan sphere.*

4.2.3 Molecular Dynamic simulation (MM/QM)

For the molecular dynamic simulation experiments on EWL, E. Dumont & N. Gillet used a similar molecular dynamics simulations (MD) approach to AdkA in which TbXo4 intensity binding rate with each amino acids in EWL structure were captured. Blind evaluations were done using EWL structures from the PDB (HEWL: 5F14 ; TEWL: 1JEF ; QEWL: 2IHL and STEWL: 2VG0 from PDB-redo) except for PEWL and OEWL, which they use PEWL structure that I solved, as no structure was available then, and OEWL structure predicted by Swiss-Model from HEWL structure (PDB:6F2I) . The sequence alignment of different EWL used in MD simulation, with HEWL as reference, are presented in Figure 4-19. The simulations were replicates of 5 times 200ns, in the presence of 10 crystallophores with one EWL molecule.

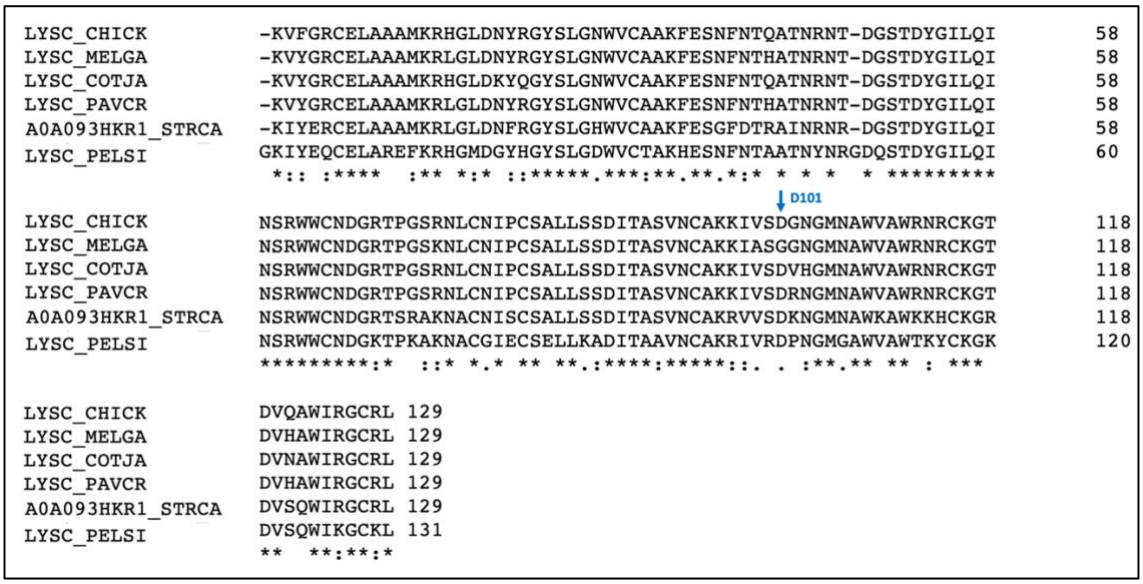


Figure 4-19 *Multiple sequence alignment by Clustal Omega of Lysozymes from Hen Egg-White (LYSC_CHICK), Turkey Egg-White (LYSC_MELGA), Quail Egg-White (LYSC_COTJA), Peacock Egg-White (LYSC_PAVCR) Ostrich Egg-White (STRCA) and Softshell Turtle Egg-White (LYSC_PELSI)*

The results from the MD simulations are shown in Figure 4-20. Noted that we added two other EWL sequences from Ostrich and Shell turtle. As mentioned above, OEW does not express the desired lysozyme. As for STEW, Shell turtle express Lysozyme C. However, it is a protected species so that we do not easily have access to eggs.

Also, The choice to add these two EWL to the simulation is due to the presence of close similarity between OEWL and PEWL as OEWL present Lysine at the residue 102 which is Arginine in PEWL. As for STEWL, it presents two interesting mutations around D101, S100R and G102P. Hence we took advantage of the PDB and MD simulations which has allowed to expand our evaluation of mutations on OEWL and STEWL.

The results from the MD simulations confirm that D101 interact with the TbXo4 based on the interaction intensity rate of TbXo4 with the residue in EWL that possess D101, except for STEWL where no interaction with TbXo4 where observed near D101,which was replaced by a salt bridge interaction between R97 and D101, and also for TEWL which present a mutation at D101G . Also, it has confirmed the important role of W62 and W63 as the interaction with these two amino acids is present in all EWL, even when D101 is absent. In addition, other sites were detected, such as D119 in TEWL and QEWL. For PEWL, the simulations had determined that PEWL presents a dimer of TbXo4 near D101, with R102 in PEWL interacts with the carboxylate of TbXo4.
In the case of OEWL, the simulations reveal two distinct positions of the loop comprising residues 101-103. While the loop open position allows the interaction of a dimer of TbXo4, the close one only shows the interaction of one TbXo4 molecule.
Finally, for STEWL, the MD simulations indicate that TbXo4 still interact with the protein surface but the interaction does not involve D103 residue (corresponding to D101 in HEWL), as the later residue forms a salt bridge with R99. However, TbXo4 still interacts with W64-W65 (as W62-W63 in HEWL).

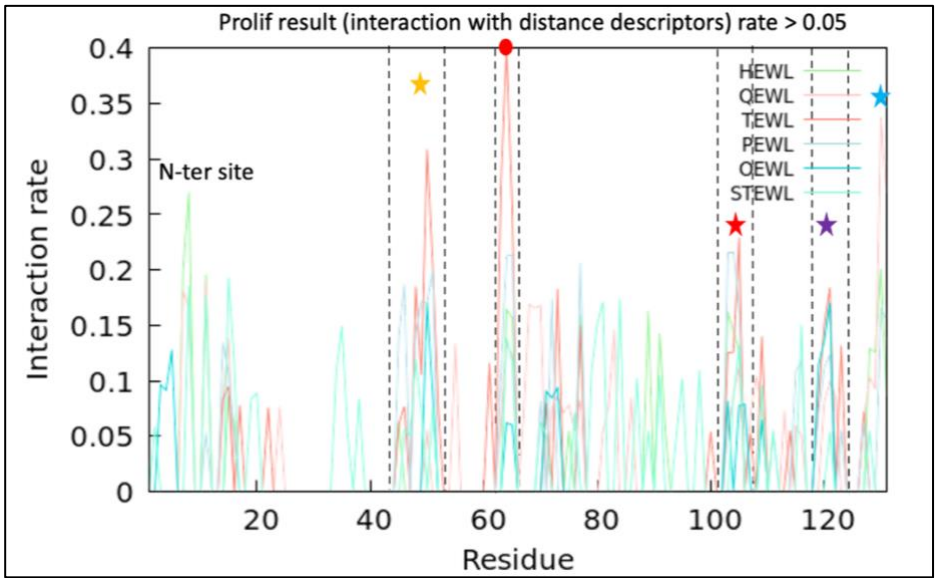


Figure 4-20 *Interaction rates of TbXo4 with EWL residues during MM/QM molecular dynamics simulation of 5 times 200ns duration.*

4.2.4 Conclusion EWL

From EWL crystallization assays and the crystallization diagrams, TEWL has shown the importance of D101 for the nucleating effect of TbXo4 in EWL. Indeed, by interacting with it, TbXo4 stabilises D101 and reduces the surface entropy associated with, leading to induce crystallization hits. Also, from QEWL and PEWL analysis, we have shown the importance of the nature of amino acids present in the vicinity of TbXo4 interaction for EWL crystallizability. Indeed the QEWL structure has shown a more stable TbXo4 complex compared to the situation in HEWL structure and is due to the mutations of G102 and N103 to a bulkier and charged amino acids V and H, respectively, stabilizing the picolinate cycle of the complex. However, as revealed in PEWL structure the mutation of G102 to a positively charged such as R creates a steric hindrance inducing a change in TbXo4 conformation and therefore preventing the crystallisation of PEWL.

The MD simulation results have confirmed the main amino acids interacting with TbXo4, such as D101 in HEWL and QEWL, W62 and W63 in all EWL, but have also suggested other residues to be considered such as D119 and G102.

Therefore, it would be interesting in the future to carry out a study of the effects of other mutations in the interaction sites (like W62-W63, N103, W123 or D119) by expressing the protein in yeast or by cell free, as it may be difficult to get natural proteins with the desired mutations.

CHAPTER 5

CONCLUSION AND PERSPECTIVES

5. CONCLUSION AND PERSPECTIVES

Over the last two decades, lanthanide complexes have been proposed as phasing agents in order to overcome the phase problem thanks to the significant anomalous contribution of lanthanide ions. Historically, three generations of neutral and negatively charged lanthanide complexes have been initially exploited and studied and led to the development of a stable and mono-cationic lanthanide complex TbXo4 or crystallophore with phasing properties and more importantly a strong nucleating effect. Highlighting TbXo4 exceptional properties was at the heart of the PhD work of Sylvain Engilberge. Despite the important properties that TbXo4 provides, the different studies undergone on it show that it is not 100% successful, in particular in term of crystallization. Hence increasing its success rate is of importance and implies a better understanding of the interaction and nucleation properties of TbXo4 and its behaviour on the protein surface. This was the starting point of the present work entitled "Developments of Lanthanide complexes for structural biology ".

During this thesis, we have addressed the question by a two-fold approach. The first side of the approach has consisted in exploring the chemistry offered by lanthanide complexes by evaluating the nucleating properties of thirteen crystallophore variants, provided by our collaborators at ENS Lyon, on a panel of three commercial proteins. This approach has enabled us to select a new variant of Xo4, TbXo4-OH, characterized by a propanol extension on the free nitrogen of the triazacyclononane macrocycle. TbXo4-OH showed good nucleating efficiency as it provides a high number of crystallization hits as well as additional hits when compared to the reference molecule TbXo4. Also, this variant seems to bind better at the surface of the protein, due to its steric hindrance, compared to the other variants evaluated. In addition, molecular dynamic simulation, performed by our collaborator at the ENS of Lyon, have confirmed that TbXo4-OH creates weak interactions while exploring the proteins surface, due to its alcohol extension. This work has highlighted the importance of ligand dynamics and conformations on the nucleating properties of crystallophore. Of course, it would be interesting, in the future, to evaluate Xo4 variants on a larger panel of proteins, in particular by exploiting some unique conditions relative to each variants in order to better understand the influence of the variants' moieties on their binding modes. Finally, the TbXo4-OH variant has been selected as the second generation of

crystallophore and has been made available to the structural biology community through the Polyvalan company who already provide TbXo4. Moreover, throughout this thesis, we have also encountered experimental problems with the precipitation of some proteins during mixing with TbXo4, while trying to reproduce the protocols used in previous works. This difficulty was at the end a benefit, as it has shown that the crystallophore obtained from a new production and purification protocol is more efficient compared to the initially produced one and induces in some cases crystalline precipitates which can be used for MMS experiment, for example. As a consequence the 1+1+1 crystallization approach evaluated during this thesis to control the increased nucleating properties of TbXo4 produced thanks to the improved purification process is now the new protocol of use of the molecule as indicated on Polyvalan's website (https://crystallophore.fr/nucleating-agent/).

The second approach was a biological approach that has consisted in analysing the interaction sites between TbXo4 and the protein target. This analysis was combined with mutagenesis (point mutation) in order to confirm amino acids or groups of residues that are related in the binding of the molecule. This approach was applied on two protein models (MthAdkA and EWL). The results gained on the first model, MthAdkA were not as informative as expected. Despite several generated mutants, we did not succeed in suppressing the nucleating effect of TbXo4 on this protein. Even though one can consider that this protein was not the best model to study the interaction of TbXo4 in relation with its nucleating outputs, the performed study has highlighted the role of minor binding sites in the nucleating properties of the crystallophore, that have to be identified and taken into account in future studies. For the second model (EWL), it has been validated that D101, which directly interact with TbXo4, plays an important and main role in the nucleating effect of TbXo4 on EWL. Moreover, our study has also shown that amino acids close to position 101 influence the interaction of TbXo4 with EWL surfaces, as highlighted by the mutation at position 102. Again, while a direct interaction between TbXo4 lanthanide ion with negatively charged amino acid (Asp and Glu) seems to be the most favourable situation to promote its nucleating effect, the nature of the residues surrounding the main interacting residue modulate TbXo4 crystallization properties. Even though we are still far to be able to predict a "required" cluster for the interaction of TbXo4 with a protein target in order to promote the nucleation of the latter, the results from EWL are

promising and require to be extended. In particular, it would be interesting to apply the mutagenesis approach on another protein model, such as Glyoxylate reductase/hydroxypyruvate reductase (GRHPR). Indeed, this protein was in the panel used to highlight TbXo4 properties. While it presents a high percentage of negatively charged amino acids on its surface, it only shows low crystallization hits in the presence of TbXo4.

Last but not least, the present work has demonstrated the complementarity between molecular dynamics and crystal structure analysis. While MD allows to figure out the very first steps of TbXo4 interaction and thus nucleation, the crystal structure provides the very last image of the crystallization process. As a consequence, we are still missing the steps that occur in solution, ideally with thermodynamic values associated to the nucleating properties of TbXo4. During this thesis, I ran preliminary experiments, such as Isothermal Titration Calorimetry, Dynamic/Static Light Scattering and nano Differential Scanning Fluorimetry. If I was able to get differences in signal when TbXo4 is present or not, we have so far not been able to correlate these differences with the crystallization results we observed, leaving room for future studies.

CRYSTALLOPHORE, NOUVELLE REVOLUTION EN BIOLOGIE STRUCTURALE

RESUME DE LA THESE EN FRANÇAIS

6. RESUME DE LA THESE EN FRANÇAIS

6.1 Introduction

La cristallographie aux rayons X est considérée comme la méthode la plus productive et la plus puissante pour obtenir des informations structurelles de haute précision sur les macromolécules biologiques, puisque 86,17 % des structures de macromolécules biologiques déposées dans la banque de données Protein Data Bank (PDB) ont été déterminées par cette méthode (RSCB Protein Data Bank Novembre-2022).

Aujourd'hui, la cristallographie macromoléculaire (MX) est concurrencée par la cryo-EM pour la détermination des structures de bio-macromolécules (Subramaniam, 2019), car cette dernière a connu de nombreux développements qui en font une alternative efficace à la cristallographie aux rayons X.

Malgré les progrès importants connues par les autres méthodes structurales, la cristallographie aux rayons X reste une approche de choix pour obtenir des données structurales à l'échelle atomique des macromolécules biologiques. En effet, la cristallographie a grandement participé à la compréhension de la fonction des protéines et à l'élucidation de leurs mécanismes facilitant, par exemple, le développement de nouveaux médicaments basés sur une conception rationnelle. Malgré tout, la cristallographie souffre encore de trois problèmes principaux :

1) l'obtention d'une quantité de protéine suffisante pour effectuer la cristallisation, cette limitation pouvant être surmontée par la maîtrise de l'expression et purification de la protéine,

2) la production de cristaux de grande qualité et présentant une bonne diffraction et

3) la résolution du problème des phases inhérent à cette méthode et pour lequel plusieurs techniques ont été développées pour faciliter sa résolution telles que la méthode du Remplacement moléculaire (RM), ou les méthodes exploitant la diffusion anormale (SIR/MIR, SAD/MAD et SIRAS/MIRAS). Plus récemment, des développements exploitant de l'intelligence artificiel ont rendu la résolution du problème des phases moins problématique puisque des programmes tels que Alphafold (Jumper et al., 2021; Varadi et al., 2022) et RoseTTAFold (Baek et al., 2021b)

permettent de prédire le modèle 3D d'une protéine à partir de sa séquence en acides aminés, modèle qui peut par la suite être exploité pour le RM (Barbarin-Bocahu & Graille, 2022).

Afin de dépasser les deux derniers obstacles, des développements méthodologiques et instrumentaux ont été mis en place. Par exemple, nous pouvons citer :

1) L'automatisation de la cristallisation combinée à l'utilisation de volumes faibles de matériel biologique et au criblage de dizaines/centaines conditions physico-chimiques grâce à de nombreux kits de cristallisation disponibles sur le marché (Brown et al., 2003; Santarsiero et al., 2002).

2) La nano-cristallographie utilisant aussi bien des sources de lumière synchrotron avec des faisceaux de taille micro/nano-métrique accordables associé à l'automatisation des lignes de lumière ; que des sources du type laser à électrons libres (XFEL) (Emma et al., 2010).

3) Le développement de détecteurs permettant une lecture rapide des données de diffraction et l'élimination du bruit de partage de charge, l'amélioration de la taille des pixels et le filtrage à faible énergie (Hülsen et al., 2006).

Ces améliorations ont ainsi permis d'autres avancées telles que les développements de la découverte de médicaments à l'aide d'approches de criblage de fragments (Blundell et al., 2002) ou la cristallographie résolue dans le temps (Mehrabi et al., 2019).

Cependant, l'analyse statistique effectuée sur les résultats du Structural Genomics Consortium (Khurshid et al., 2014) montre que seulement 10% des protéines purifiées ont vu leur structure déterminée par cristallographie. Ainsi, toutes les innovations mentionnées ci-dessus ne sont pas suffisantes et il demeure donc une marge importante de développements potentiels pour la cristallographie, en particulier dans le domaine de la cristallisation.

Par conséquent toutes approches permettant de contrôler les paramètres qui influencent la nucléation sont un plus. La rationalisation des conditions de cristallisation à travers les kits commerciaux a partiellement permis de contrôler l'influence de la nature et concentration des agents précipitants, le pH, les forces ioniques etc… mais cela reste insuffisant. De nouvelles techniques de cristallisation ont été proposées dont la meilleure illustration est l'approche de recherche de nouvelles conditions de cristallisation par ensemencement (rMMS, random microseed matrix screening, (D'Arcy et al., 2007, 2014)). Toutefois, cette dernière approche

suppose une condition permettant de constituer le stock de cristaux servant à l'ensemencement. Lorsqu'aucune condition de cristallisation ne peut être déterminée, des additifs peuvent être introduits afin de promouvoir la nucléation. On parle alors d'agents nucléant. Différentes types ont été évalués tels que des additifs naturels (crin de cheval) (D'Arcy et al., 2003), des films protéiques (Pechkova & Nicolini, 2002), des matériaux poreux (Chayen et al., 2001, 2006; Khurshid et al., 2014), ou des polymères préalablement marqués par la protéine d'intérêt (E. Saridakis et al., 2011). Enfin, plus récemment des molécules agissant comme des « glues moléculaires » ont été proposées et comportent principalement trois familles i.e. les Polyoxométalates (Bijelic & Rompel, 2015, 2017), les Calixarènes (Alex et al., 2018; Shinkai et al., 1988) et les complexes de lanthanides sur lesquels nous travaillons (Engilberge et al., 2017). Dans ce contexte, au cours des deux dernières décennies, les complexes de lanthanides ont été étudiés par notre groupe, initialement comme outil pour résoudre le problème de la phase grâce, notamment à la contribution anomale significative des ions lanthanide (Djinovic-Carugo & Carugo, 2015). Ainsi, trois générations de complexes de lanthanides ont été développées.

La première génération qui comprend des complexes utilisant des ligands macrocycliques (DOTA, DO3A et HPDO3A) ou des ligands linéaires (DTPA, DTPA-BMA) ont été étudiés lors de la thèse de mon directeur de thèse Eric GIRARD (1998-2001) puis au cours de la thèse de Meike STELTER (2002-2005) (Girard et al., 2003b; Stelter et al., 2014). Ces complexes ont été évalués sur différentes protéines commerciales et ont permis de déterminer la structure de nouvelles protéines (Arnoux et al., 2009; Chaudhuri et al., 2003; de Bono et al., 2005; Eichmann et al., 2016; Larivière et al., 2012). Ensuite la deuxième génération de complexes de lanthanides basés sur le tris-dipicolinate (DPA) a été évaluée par Guillaume POMPIDOR (2004-2007), lors de sa thèse, où il a montré que cette famille de complexes de lanthanides qui possèdent des propriétés de luminescence, présentent aussi des effets positifs sur la nucléation et la cristallisation (Pompidor et al., 2008), tout en permettant de résoudre le problème des phases (Pompidor et al., 2010). Cette thèse a été suivie par celle de Romain TALON (2008 - 2012), qui a travaillé sur des molécules dérivées des complexes DPA et qui a étudié l'influence de la fonctionnalisation du ligand sur la cristallisation ainsi que sur le phasage de nombreuses protéines modèles (Talon et al., 2011, 2012). Cependant, ces deux générations de complexes, chargés négativement ou neutre, ne couvrent pas une large gamme de protéines et ont montré,

pour certains, des problèmes de stabilité dans les solutions de cristallisation (phénomène d'auto-cristallisation). Il était donc nécessaire de développer un nouveau complexe pour surmonter ces inconvénients. Sur la base de ces études antérieures, une troisième génération de molécules à base de lanthanide, nommée par la suite " Crystallophore ", a été conçue en collaboration avec l'équipe d'Olivier Maury de l'ENS de Lyon.

Ce complexe cationique est formé d'un ion terbium (3+) entouré d'un ligand macrocyclique triazacyclononane bis-picolinate et possède une excellente solubilité et stabilité dans les conditions physico-chimiques rencontrées dans les milieux de cristallisation. Il a été étudié par Sylvain ENGILBERGE (2014-2017) qui a révélé ses propriétés uniques. En effet, le TbXo4 possède non seulement des propriétés de phasage, lié à l'atome de terbium, mais il favorise également la cristallisation de manière significative (Engilberge et al., 2017). Il s'agit donc d'un outil tout-en-un qui permet de surmonter les deux principaux problèmes de la cristallographie macromoléculaire.

Huit protéines ont été initialement utilisées pour démontrer les capacités de TbXo4 (Engilberge et al., 2017). En particulier, il a été démontré que TbXo4 induisait de nouvelles conditions de cristallisation uniques et a été utilisé pour résoudre la structure de deux protéines inconnues (Engilberge et al., 2017). Il a été également démontré l'efficacité de l'approche TbXo4 pour la détermination de la structure, en particulier pour surmonter les problèmes souvent rencontrés tels que les cristaux maclés ou présentant une faible diffraction, pour promouvoir différents empilements de cristaux et enfin, il a montré la compatibilité de TbXo4 avec les approches de cristallographie sérielle (Engilberge et al., 2019).

Afin d'expliquer et espérer améliorer les propriétés des crystallophores, Sylvain avait également commencé à décrire l'interaction de TbXo4 telle qu'observée dans différentes structures cristallographiques (Engilberge et al., 2018). Les analyses des interactions ont été associées à des calculs basés sur la théorie de la fonctionnelle de la densité (DFT), réalisés par Elise Dumont (chimiste théoricien à l'ENS de Lyon), qui ont permis de chiffrer les énergies régissant les interactions entre le complexe et la surface de la protéine. Cette analyse a montré que TbXo4 possède différents modes de liaison, résultant de différents équilibres entre interactions électrostatiques, liaisons hydrogène ou interactions hydrophobes (Engilberge et al., 2018).

Malgré toutes les propriétés avantageuses décrites précédemment, les différentes études menées sur TbXo4 ont montré qu'il n'est pas efficace à 100% en terme de cristallisation, laissant un espace pour des améliorations.

Ainsi déchiffrer l'origine des propriétés de nucléation du TbXo4 et son mécanisme d'interaction sur la surface des protéines est important afin d'augmenter son taux de réussite.

Pour cela, notre approche a été double :

1) La première approche est chimique. Elle consiste à exploiter la diversité chimique offerte par les complexes de lanthanides, en particulier TbXo4. A cet égard, plusieurs molécules dérivées de TbXo4 (des variants) ont été synthétisées par l'ajout de fonctions chimiques soit directement sur le macrocycle, soit sur le bras picolinate. Afin d'évaluer l'influence de ces ajouts sur la cristallisation, les propriétés de nucléation des variants ont été déterminées en utilisant la plateforme de cristallisation HTX Lab (https://embl.fr/htxlab/) sur un panel de 3 protéines commerciales. Cette approche (Chapitre 3) a apporté un nouvel éclairage sur la question complexe de la cristallisation des protéines mais a également permis d'optimiser l'utilisation du TbXo4.
2) La seconde approche est une approche biologique qui consiste à analyser les sites d'interaction entre les surfaces des protéines et le TbXo4, par analyse de la structure cristalline combinée à des calculs théoriques. Cette approche est complétée par une approche de mutagenèse (mutations ponctuelles) pour cibler les motifs protéiques potentiellement impliqués dans la liaison de TbXo4. Cette seconde approche (Chapitre 4) a permis de déterminer le rôle prédominant de certains acides aminés dans la cristallisation induite par TbXo4.

Dans les parties suivantes, je vais décrire brièvement chacune des approches présentées ci-dessus ainsi que les résultats obtenus, au cours de cette thèse, pour chaque approche.

6.2 Évaluation des variants de TbXo4

Dans le but de comprendre le mécanisme de nucléation du TbXo4 et son comportement à la surface des protéines et afin d'augmenter son taux de réussite,

cette première approche explorée au cours de cette thèse consiste à exploiter la chimie offerte par les complexes de lanthanides, en particulier celle du ligand Xo4, pour déchiffrer l'influence potentielle de différentes fonctions chimiques sur les propriétés de nucléation du crystallophore. Pour cela, nous avons étudié une série de 13 molécules synthétisées par nos collaborateurs de l'ENS-Lyon, soient directement basées sur Xo4, soient sur des molécules dérivées.

Tout d'abord, pendant cette étude en particulier, en essayant de reproduire le protocole de préparation des échantillons biologiques utilisé par S. Engilberge (Engilberge et al., 2017), nous avons rencontré des problèmes de précipitation des protéines, non observés précédemment. Cependant, ce problème de précipitation n'en est finalement pas un puisqu'il est associé à un effet nucléant augmenté du TbXo4 du fait d'un nouveau protocole de synthèse/purification amélioré de la molécule. Par conséquent, nous avons proposé un nouveau protocole d'utilisation de TbXo4 afin d'éviter la précipitation des protéines. Ce protocole, dit 1+1+1, consiste à mélanger, dans cet ordre, 1 volume de protéine + 1 volume de TbXo4 suivie par 1 volume de la condition de cristallisation. Compte tenu de sa mise en place tardive, ce protocole n'a pas été utilisé pour les expériences de cristallisation automatisées, pour lesquelles un protocole du type 1 volume protéine/TbXo4 + 1 volume de la condition de cristallisation est exploité, mais a été appliqué pour les cristallisations manuelles.

Nous avons donc évalué les propriétés de nucléation des treize molécules dérivées de TbXo4 sur trois protéines commerciales (HEWL, ProteinaseK et Thaumatine) et nous les avons comparé à celle de TbXo4. Ceci nous a permis de déterminer que le variant nommé TbXo4-OH, qui présente un bras propanol sur l'azote libre du macrocycle triazacyclononane, présente les meilleurs résultats parmi les treize variants sur ces trois protéines en fournissant à la fois un taux élevé d'occurrences de cristallisation et à la fois des occurrences complémentaires à celle de la molécule de référence TbXo4.

Ensuite en exploitant des données de diffraction obtenues sur le lysozyme de blanc d'œuf de poule (HEWL), comme protéine modèle, nous avons décrit le mode de liaison des variants les plus efficaces et nous les avons comparé à celui observé pour le TbXo4, en présence de ces composés. L'analyse des structures de HEWL issues de cristaux obtenus en présence de 10 mM de ces variants a montré que le variant TbXo4-OH, choisi sur la base des résultats de cristallisation, montre que cette molécule présente une conformation différente du ligand à la surface de la protéine,

comparée à celle de TbXo4. Il est capable d'interagir avec différents types d'acides aminés, en générant des interactions hydrophobes (CH-π) via sa partie macrocycle (TACN), d'interagir avec des acides aminés chargés positivement tels que l'arginine par une de ses fonctions carboxylate présente sur la partie picolinate. De plus, ce variant semble mieux se lier à la surface de HEWL. Ceci est lié à son encombrement stérique, par rapport aux autres variantes. Ces observations ont été confirmées par simulation de dynamique moléculaire de HEWL en présence des mêmes variants et ont été réalisées par nos collaborateurs à l'ENS-Lyon, Elise DUMONT et Natacha GILLET. Les simulations montrent que TbXo4-OH, en raison de son groupement alcool (rôle de donneur/accepteur), crée de faibles interactions lors de l'exploration de la surface des protéines, car il présente une flexibilité plus élevée par rapport aux autres variants.

Pour ces raisons, nous proposons que ce variant constitue la seconde génération de crystallophore. Toutefois, ces résultats préliminaires devront être consolidés par une évaluation de chacun des variants sur un panel plus large de protéines, que j'ai partiellement initié, au cours de cette thèse, sur un ensemble de 8 protéines commerciales, sélectionnées pour couvrir différents poids moléculaires et assemblages ainsi que différentes valeurs de pI. De plus, il serait intéressant d'exploiter les conditions uniques de cristallisation offertes par chacun des variants afin de mieux décrire l'influence de chaque groupement chimique sur la liaison de ces molécules à la surface des protéines.

6.3 Caractérisation et compréhension du mécanisme d'interaction de TbXo4

Concernant la seconde approche exploitée au cours de cette thèse, elle est basée sur une étude qui a montré qu'il était possible de réaliser des mutations dans le domaine conservé d'un nano-body afin de permettre à des fluorochromes ou des métaux lourds de s'y fixer (Hansen et al., 2017). De manière similaire, nous aimerions effectuer des modifications par mutagenèse dirigée permettant la création d'un site de liaison TbXo4 favorisant l'interaction du complexe avec une protéine d'intérêt, favorisant par la suite sa cristallisation.

Ainsi, il est donc important d'analyser l'interaction de TbXo4 avec les protéines afin de déterminer le rôle prédominant de certains acides aminés dans la cristallisation induite par TbXo4. Nous avons utilisé une approche de mutagenèse (mutation ponctuelle)

pour cibler les motifs protéiques potentiellement impliqués dans la liaison du complexe et confirmer ou non le rôle de certains acides aminés.
Pour cela, nous avons choisi deux protéines modèles pour cette étude :
1) Le premier modèle de protéine est l'Adénylate Kinase de *Methanothermococcus thermolithotrophicus* qui forme un assemblage trimérique et présente un mode de fixation original ainsi qu'une asymétrie de fixation de TbXo4 (Engilberge et al., 2019; Roux et al., 2021), comme décrit dans le paragraphe 6.3.1.
2) Le lysozyme de l'œuf blanc de poule (HEWL), une protéine commerciale fournie par Roche (référence : 10837059001), qui possède un grand nombre de conditions uniques de cristallisation induites par TbXo4 lors des criblages au laboratoire HTX (159 occurrences uniques) (Engilberge et al., 2017).

Dans cacun des deux cas, j'ai utilisé la plateforme HTX-Lab afin de suivre le taux de cristallisation de chaque mutant de notre protéine d'intérêt (en présence et en absence de TbXo4) et j'ai aussi effectué des cristallisations manuelles dans une condition conventionnelle pour les comparer aux résultats obtenus avec la protéine de type sauvage.
Enfin, des simulations de dynamique moléculaire ont été réalisées par nos collaboratrices de l'ENS-Lyon (Elise Dumont et Natacha Gillet) pour analyser nos deux protéines modèles et leurs mutants, afin de déterminer les différents modes de liaison de TbXo4 et les énergies régissant les interactions entre le complexe et la surface des protéine/mutants considérés.

Les deux parties suivantes décrivent pour chaque modèle de protéine les expériences que nous avons réalisé ainsi que les résultats obtenus.

6.3.1 Adénylate kinase

Les adénylate kinases (AKs, ATP/AMP phosphotransférase, EC2.7.3.4) sont de petites enzymes (~20 KDa) qui maintiennent l'équilibre des niveaux d'adénylate cellulaire en catalysant l'interconversion des nucléotides d'adénine, et jouent donc un rôle important dans l'homéostasie énergétique cellulaire en suivant la réaction ci-dessous :

$$Mg^{2+}\ ATP + AMP \Leftrightarrow Mg^{2+}ADP + ADP$$

Tristan Wagner (Max Planck Institute, Bremen), collaborateur de ce projet depuis 2016 et intéressé par l'étude du mécanisme ci-dessus, avait réalisé une purification native de l'adénylate kinase (AdkA) de l'archée marine anaérobie *Methanothermococcus thermolithotrophicus* (Engilberge et al., 2019; Vögeli et al., 2018). Grâce à l'utilisation du TbXo4, des données de diffraction de MthAdkA ont été obtenues à partir d'un cristal de la protéine co-cristallisé en présence de 10 mM de TbXo4, puis trempé dans une solution cryoprotectrice contenant 50 mM de TbXo4 (Engilberge et al., 2019). La structure a révélé un mode de liaison original de TbXo4 (Roux et al., 2021).
En effet, les données de diffraction ont permis d'identifier la présence d'un assemblage trimérique dans l'unité asymétrique (3 monomères d'AdkA : A, B et C) correspondant à l'unité biologique et d'identifier un mode de fixation original avec une asymétrie de liaison (Engilberge et al., 2019; Roux et al., 2021):

1) Deux sites de liaison forts ont été identifiés sur les monomères A et B. Les ions terbium interagissent avec l'Aspartate D90 et le site de fixation implique aussi la présence de Mg2+ et d'une molécule de glycérol. Cependant, aucun site de liaison équivalent sur le monomère C n'a été détecté. En effet, ce site de fixation n'est pas aussi accessible au solvant du fait de l'empilement cristallin, ce qui explique cette asymétrie de liaison. Ainsi les deux sites de liaison forts résulteraient de l'étape de trempage et non du processus de co-cristallisation.
2) Deux autres sites de liaison avec de faibles taux occupations pour le Tb ont été trouvés: un sur le monomère A (près du glutamate E50) et que l'on observe pas sur les autres monomères et qui pourrait être « une trace » du processus de cristallisation induit par TbXo4. Le second site de faible liaison n'a été observé que sur le monomère C. Il n'est pas clair si ce site résulte du processus de cristallisation ou de l'étape de trempage.

Ces observations ont montré la nécessité de recueillir des données de diffraction d'AdkA à partir d'un cristal co-cristallisé avec 10mM TbXo4 sans trempage.

Ainsi, pour confirmer nos hypothèses, nous avons collecté des données de diffraction sur un cristal de la protéine recombinante (MthAdkA WT) co-cristallisé en présence de 10 mM de TbXo4. Nous avons, tout d'abord, confirmé que MthAdkA WT se comporte de manière similaire à la protéine produite de manière native, notamment en terme de cristallisation en présence de TbXo4.

Ensuite, dans le but de déchiffrer le mécanisme de cristallisation de MthAdkA induit par TbXo4, j'ai exprimé et purifié deux séries de mutants de MthAdkA :

a) La première série repose sur les observations obtenues de la structure native. Les deux mutants présents dans cette série correspondent à la mutation du résidu E50 (Mutant 1 ou M1) et du résidu D90 (Mutant 2 ou M2) par un acide aminé Alanine.
b) La seconde série de mutations est principalement basée sur les résultats des simulations de dynamique moléculaire d'AdkA réalisées par nos collaborateurs de l'ENS-Lyon (E. Dumont & N. Gillet). En bref, la liaison de TbXo4 à l'AdkA trimérique a été capturée par dynamique moléculaire, avec un échantillonnage étendu de 10 réplicats de 200ns. Les résultats des simulations MD révèlent 6 points chauds présentant des temps de contact élevé du TbXo4. Les résidus impliqués sont : E29, E45, E49, E54, E136 et E185, respectivement. A partir des résultats de la simulation, nous avons donc choisi de nous concentrer initialement sur les résidus 54 et 136 car ils présentent les valeurs de contact les plus élevées. Nous avons également décidé de modifier notre stratégie de mutation pour limiter les problèmes potentiels de production des mutants et, par conséquent, nous avons substitué le glutamate en glutamine et, dans le cas du résidu 54, nous avons également évalué le rôle potentiel de la longueur de la chaîne latérale en le mutant en aspartate. En résumé, les mutations évaluées sont E54Q (mutant 3 ou M3), E54D (mutant 4 ou M4) et E136Q (mutant 5 ou M5).

Après purification, tous les échantillons ont été concentrés à 10 mg/mL avec ajout de 10 % de glycérol, afin d'être dans les mêmes conditions que la protéine native. Ensuite, tous les mutants ainsi que la MthAdkA WT (comme référence) ont été évalués sur la plateforme HTX (596 conditions) De plus des diagrammes de cristallisation dans les conditions conventionnelles de MthAdkA (utilisant le PEP629 comme agent précipitant) ont été réalisés pour chaque mutant. Les résultats de cristallisation des mutants ont été comparés à ceux de la MthAdkA WT.
Par la suite, nous avons exploité plusieurs cristaux pour MthAdkA WT et les mutants co-cristallisés avec 10 mM TbXo4 et nous avons confirmé l'absence ou la présence

de sites de liaison à TbXo4, pour chacun des mutants, en exploitant la contribution anomale de TbXo4.
Enfin, sur la base des résultats préliminaires obtenus sur M1 à M5, nous avons décidé de concevoir un hexa-mutant incluant les mutations suivantes : E54Q, E49Q, E136Q, E185Q, E29Q et E149Q (Mutant 6 ou M6). Le choix de muter E149 en Q a été fait car plusieurs synthèses de Fourier anomale présentaient des pics proches de E149.

Sur la base des diagrammes de cristallisation, nous pouvons dire que M1 et M2 sont des mutants intéressants pour MthAdkA. Les deux mutants M1 et M2 ont montré que les résidus E50 et D90 sont importants pour les interactions avec TbXo4 et pour induire la nucléation dans la condition de cristallisation basée sur l'agent précipitant PEP. Cependant, à partir de l'essai de cristallisation HTX, les deux mutations n'ont pas affecté de manière significative l'action de TbXo4 car les nombres de conditions de cristallisation totales et uniques de TbXo4 sont peu modifiés. Par conséquent, M1 et M2 n'offrent pas beaucoup d'informations pour comprendre l'interaction de TbXo4 dans MthAdkA.
A l'inverse, les mutations associées aux mutants M3 et M4 ont montré qu'elles avaient un effet significatif sur la cristallisation de MthAdkA. En effet, ces deux mutants ont montré une influence importante de la mutation sur la cristallisabilité de la protéine native, dans le cas de M4, ainsi que l'influence de la mutation de E54 sur la cristallisation de MthAdkA en présence de TbXo4. De plus, à partir des données obtenues, il a également été montré le rôle d'acides aminés spécifiques tels que E49, E136 et E185, en complément de E54, ainsi que l'influence des conditions de cristallisation sur la fixation de TbXo4.
Malgré plusieurs mutants générés, nous n'avons pas réussi à supprimer totalement l'effet nucléant de TbXo4 sur cette protéine, qui présente de nombreux acides aminés chargés négativement (Asp et Glu) à sa surface. Même si l'on peut considérer que cette protéine n'est pas le meilleur modèle pour étudier l'interaction de TbXo4 en relation avec ses effets nucléant, l'étude réalisée a mis en évidence le rôle des sites de fixation mineurs dans les propriétés nucléantes du crystallophore, qui doivent donc être identifiés et prises en compte dans les futures études.

6.3.2 Chasse aux œufs : Lysozyme de blanc d'œuf (EWL)

Le second modèle largement utilisé pour la caractérisation des interactions du crystallophore avec les protéines est le lysozyme de blanc d'œuf de poule (HEWL). HEWL est une protéine commerciale qui possède un grand nombre de conditions de cristallisation uniques induites par le TbXo4 pendant les criblages du HTX-Lab (Engilberge et al., 2017), et qui a également été utilisé pour évaluer les propriétés de nucléation des variantes de crystallophore, comme présenté précédemment.

Comme nous l'avons vu au chapitre 3, la structure de HEWL avec 10 mM de TbXo4 montre que le résidu D101 est le principal résidu dans l'interaction du crystallophore avec EWL par l'intermédiaire d'une coordination directe du carboxylate avec l'ion terbium du crystallophore. Par conséquent, il est intéressant d'étudier l'effet de la mutation de D101 sur la nucléation du lysozyme en présence de TbXo4.

Malheureusement, en raison de la fonction antimicrobienne du lysozyme qui consiste en l'hydrolyse des composants peptidoglycanes des parois cellulaires bactériennes, l'expression des lysozymes dans *E. coli* n'est pas facilement réalisable car elle entraîne une lyse rapide des cellules pendant son expression, ce qui conduit à un faible rendement de la protéine (Fischer et al., 1993).

Par chance, nous avons découvert que la mutation D101G existe naturellement dans le lysozyme blanc d'œuf de dinde (TEWL). En conséquence, nous avons recherché d'autres lysozymes susceptibles de présenter des mutations appropriées dans le cluster qui lie TbXo4 (c'est-à-dire à proximité de D101 ou d'autres résidus impliqués dans les interactions secondaires) et pour lesquels la disponibilité des œufs peut être gérée. Nous avons trouvé le lysozyme blanc d'œuf de caille (QEWL) qui présente naturellement les mutations G102V et N103H, qui sont des résidus plus volumineux et à proximité de D101, le lysozyme blanc d'œuf de paon (PEWL) qui présente la mutation G102R qui est un résidu chargé positivement à côté de D101.

En outre, nous avons également considéré deux autres lysozymes provenant du blanc d'œuf d'émeu et d'autruche (EEW et OEW) qui possèdent la mutation G102K. Cependant, il s'avère que ces deux espèces n'expriment pas l'isoforme de lysozyme souhaitée qui est le lysozyme C (C pour Chicken) ou type conventionnel. OEW et EEW exprime l'isoforme G du lysozyme (G for Goose), qui n'a pas de similarité significative avec l'isoforme C (Maehashi et al., 2012; Weaver et al., 1984). Par

conséquent, nous n'avons pas été en mesure d'exploiter expérimentalement la mutation G102K.

Par contre, cette mutation (G102K), ainsi que les mutations mentionnés ci-dessus, ont été exploitées par simulation de dynamique moléculaire (MD), réalisées par nos collaborateurs de l'ENS-Lyon (E. Dumont & N. Gillet), avec une approche similaire que celle d'AdkA. A l'exception de PEWL, ces expériences ont été réalisées en sélectionnant des structures de EWL présentes dans la PDB, pour une évaluation en aveugle, car elles ne présentent pas de TbXo4 dans leur structure. Dans le cas de PEWL, puisque qu'aucune structure n'est disponible dans la PDB, c'est la structure de PEWL que nous avons déterminée en présence de TbXo4 qui a été utilisée dans l'analyse par dynamique moléculaire.

Enfin, nous avons profité de la PDB pour tester d'autres lysozymes comme le lysozyme provenant du blanc d'oeuf de tortue molle (STEWL). Cette espèce exprime l'isoforme C du lysozyme. Cependant, il s'agit d'une espèce protégée, de sorte que nous n'avons pas facilement accès aux œufs.

La purification des trois protéines mentionnées ci-dessus à partir des œufs ainsi que la cristallisation automatisée de celles-ci ont été réalisées. Pour éviter des biais de comparaison, HEWL a été préparé dans le même tampon que ces 3 lysozymes, et une cristallisation a été réalisée au HTX-Lab en présence et en absence de 10 mM TbXo4, comme pour les 3 autres EWL. De plus, de manière similaire à l'étude AdkA, j'ai réalisé des diagrammes de cristallisation pour les différents EWL. Pour cela, j'ai exploité la condition conventionnelle de cristallisation de HEWL qui consiste en NaCl comme précipitant avec un tampon acétate de sodium à pH=4,6.

Ensuite, pour évaluer l'influence des différentes mutations naturelles sur le mode de liaison de TbXo4, j'ai déterminé les structures des trois mutants naturels à partir de cristaux co-cristallisés en présence de 10 mM de TbXo4.

Les expériences entreprises ont montré l'importance de D101 pour l'effet nucléant de TbXo4 dans les EWL. En effet, en interagissant avec ce résidu, TbXo4 réduit l'entropie de surface qui est associé à D101, favorisant ainsi la cristallisation. De plus, les analyses menées sur QEWL et PEWL ont montré l'importance de la nature des acides aminés présents à proximité de l'interaction avec TbXo4 sur la cristallisabilité des EWL. En effet, la structure QEWL a montré un complexe TbXo4 plus stable comparé

à la situation observée dans la structure HEWL. Ceci est lié aux mutations de G102 et N103 en acides aminés V et H ce qui favorise l'interaction avec le cycle picolinate du complexe. Cependant, comme le révèle la structure du PEWL, la mutation de G102 en un acide aminé chargé positivement comme R crée un obstacle stérique qui induit un changement de conformation du TbXo4 et réduit donc la cristallisation du PEWL.

Les résultats de la simulation MD ont confirmé les principaux acides aminés interagissant avec TbXo4, tels que D101 dans HEWL et QEWL, W62 et W63 dans tous les EWL, mais ont également suggéré d'autres résidus à considérer tels que D119 et G102.

Par conséquent, il serait intéressant à l'avenir de réaliser une étude des effets d'autres mutations dans les sites d'interaction (comme W62-W63, N103, W123 ou D119) en exprimant la protéine en levure ou en cellule libre, car il peut être difficile d'obtenir des protéines naturelles avec les mutations souhaitées.

REFERENCES

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). *PHENIX*: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, *66*(2), 213–221. https://doi.org/10.1107/S0907444909052925

Alex, J. M., Rennie, M. L., Engilberge, S., Lehoczki, G., Dorottya, H., Fizil, Á., Batta, G., & Crowley, P. B. (2019). Calixarene-mediated assembly of a small antifungal protein. *IUCrJ*, *6*(2), 238–247. https://doi.org/10.1107/S2052252519000411

Alex, J. M., Rennie, M. L., Volpi, S., Sansone, F., Casnati, A., & Crowley, P. B. (2018). Phosphonated Calixarene as a "Molecular Glue" for Protein Crystallization. *Crystal Growth & Design*, *18*(4), 2467–2473. https://doi.org/10.1021/acs.cgd.8b00092

Arnoux, P., Morosinotto, T., Saga, G., Bassi, R., & Pignol, D. (2009). A Structural Basis for the pH-Dependent Xanthophyll Cycle in *Arabidopsis thaliana*. *The Plant Cell*, *21*(7), 2036–2044. https://doi.org/10.1105/tpc.109.068007

Asanithi, P., Saridakis, E., Govada, L., Jurewicz, I., Brunner, E. W., Ponnusamy, R., Cleaver, J. A. S., Dalton, A. B., Chayen, N. E., & Sear, R. P. (2009). Carbon-Nanotube-Based Materials for Protein Crystallization. *ACS Applied Materials & Interfaces*, *1*(6), 1203–1210. https://doi.org/10.1021/am9000858

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., … Baker, D. (2021a). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *373*(6557), 871–876. https://doi.org/10.1126/science.abj8754

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., … Baker, D. (2021b). Accurate prediction of protein structures and

interactions using a three-track neural network. *Science*, eabj8754. https://doi.org/10.1126/science.abj8754

Barbarin-Bocahu, I., & Graille, M. (2022). The X-ray crystallography phase problem solved thanks to *AlphaFold* and *RoseTTAFold* models: A case-study report. *Acta Crystallographica Section D Structural Biology*, *78*(4), 517–531. https://doi.org/10.1107/S2059798322002157

Becker, R., & Döring, W. (1935). Kinetische Behandlung der Keimbildung in übersättigten Dämpfen. *Annalen der Physik*, *416*(8), 719–752. https://doi.org/10.1002/andp.19354160806

Bijelic, A., Molitor, C., Mauracher, S. G., Al-Oweini, R., Kortz, U., & Rompel, A. (2015). Hen Egg-White Lysozyme Crystallisation: Protein Stacking and Structure Stability Enhanced by a Tellurium(VI)-Centred Polyoxotungstate. *ChemBioChem*, *16*(2), 233–241. https://doi.org/10.1002/cbic.201402597

Bijelic, A., & Rompel, A. (2015). The use of polyoxometalates in protein crystallography – An attempt to widen a well-known bottleneck. *Coordination Chemistry Reviews*, *299*, 22–38. https://doi.org/10.1016/j.ccr.2015.03.018

Bijelic, A., & Rompel, A. (2017). Ten Good Reasons for the Use of the Tellurium-Centered Anderson–Evans Polyoxotungstate in Protein Crystallography. *Accounts of Chemical Research*, *50*(6), 1441–1448. https://doi.org/10.1021/acs.accounts.7b00109

Bijelic, A., & Rompel, A. (2018). Polyoxometalates: More than a phasing tool in protein crystallography. *ChemTexts*, *4*(3), 10. https://doi.org/10.1007/s40828-018-0064-1

Blundell, T. L., Jhoti, H., & Abell, C. (2002). High-throughput crystallography for lead discovery in drug design. *Nature Reviews Drug Discovery*, *1*(1), 45–54. https://doi.org/10.1038/nrd706

Boistelle, R., & Astier, J. P. (1988). Crystallization mechanisms in solution. *Journal of Crystal Growth*, *90*(1–3), 14–30. https://doi.org/10.1016/0022-0248(88)90294-1

Brown, J., Walter, T. S., Carter, L., Abrescia, N. G. A., Aricescu, A. R., Batuwangala, T. D., Bird, L. E., Brown, N., Chamberlain, P. P., Davis, S. J., Dubinina, E., Endicott, J., Fennelly, J. A., Gilbert, R. J. C., Harkiolaki, M., Hon, W.-C., Kimberley, F., Love, C. A., Mancini, E. J., … Stuart, D. I. (2003). A procedure for setting up high-throughput nanolitre crystallization experiments. II. Crystallization results. *Journal of Applied Crystallography*, *36*(2), 315–318. https://doi.org/10.1107/S0021889803002012

Bruins, M. E., Matser, A. M., Janssen, A. E. M., & Boom, R. M. (2007). Buffer selection for HP treatment of biomaterials and its consequences for enzyme inactivation studies. *High Pressure Research*, *27*(1), 101–107. https://doi.org/10.1080/08957950601082573

Chaudhuri, B. N., Sawaya, M. R., Kim, C.-Y., Waldo, G. S., Park, M. S., Terwilliger, T. C., & Yeates, T. O. (2003). The Crystal Structure of the First Enzyme in the Pantothenate Biosynthetic Pathway, Ketopantoate Hydroxymethyltransferase, from M. tuberculosis. *Structure*, *11*(7), 753–764. https://doi.org/10.1016/S0969-2126(03)00106-0

Chayen, N. E., Saridakis, E., El-Bahar, R., & Nemirovsky, Y. (2001). Porous silicon: An effective nucleation-inducing material for protein crystallization 1 1Edited by R. Huber. *Journal of Molecular Biology*, *312*(4), 591–595. https://doi.org/10.1006/jmbi.2001.4995

Chayen, N. E., Saridakis, E., & Sear, R. P. (2006). Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium. *Proceedings of the National Academy of Sciences*, *103*(3), 597–601. https://doi.org/10.1073/pnas.0504860102

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). *MolProbity*: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography*, *66*(1), 12–21. https://doi.org/10.1107/S0907444909042073

D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W., & Marsh, M. (2014a). Microseed matrix screening for optimization in protein crystallization: What have we learned? *Acta Crystallographica Section F Structural Biology Communications*, *70*(9), 1117–1126. https://doi.org/10.1107/S2053230X14015507

D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W., & Marsh, M. (2014b). Microseed matrix screening for optimization in protein crystallization: What have we learned? *Acta Crystallographica Section F Structural Biology Communications*, *70*(9), 1117–1126. https://doi.org/10.1107/S2053230X14015507

D'Arcy, A., Mac Sweeney, A., & Haber, A. (2003). Using natural seeding material to generate nucleation in protein crystallization experiments. *Acta Crystallographica Section D Biological Crystallography*, *59*(7), 1343–1346. https://doi.org/10.1107/S0907444903009430

D'Arcy, A., Villard, F., & Marsh, M. (2007). An automated microseed matrix-screening method for protein crystallization. *Acta Crystallographica Section D Biological Crystallography*, *63*(4), 550–554. https://doi.org/10.1107/S0907444907007652

Davies, T. G., & Tickle, I. J. (2011). Fragment Screening Using X-Ray Crystallography. In T. G. Davies & M. Hyvönen (Eds.), *Fragment-Based Drug Discovery and X-Ray Crystallography* (Vol. 317, pp. 33–59). Springer Berlin Heidelberg. https://doi.org/10.1007/128_2011_179

de Bono, S., Riechmann, L., Girard, E., Williams, R. L., & Winter, G. (2005). A segment of cold shock protein directs the folding of a combinatorial protein. *Proceedings of the National Academy of Sciences*, *102*(5), 1396–1401. https://doi.org/10.1073/pnas.0407298102

Djinovic-Carugo, K., & Carugo, O. (2015). Structural biology of the lanthanides—Mining rare earths in the Protein Data Bank. *Journal of Inorganic Biochemistry*, *143*, 69–76. https://doi.org/10.1016/j.jinorgbio.2014.12.005

Doublié, S. (2007). Production of Selenomethionyl Proteins in Prokaryotic and Eukaryotic Expression Systems. In J. M. Walker & S. Doublié (Eds.), *Macromolecular Crystallography Protocols* (Vol. 363, pp. 91–108). Humana Press. https://doi.org/10.1007/978-1-59745-209-0_5

Dupeux, F., Röwer, M., Seroul, G., Blot, D., & Márquez, J. A. (2011). A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallographica Section D Biological Crystallography*, *67*(11), 915–919. https://doi.org/10.1107/S0907444911036225

Eichmann, C., Tzitzilonis, C., Nakamura, T., Kwiatkowski, W., Maslennikov, I., Choe, S., Lipton, S. A., & Riek, R. (2016). S-Nitrosylation Induces Structural and Dynamical Changes in a Rhodanese Family Protein. *Journal of Molecular Biology*, *428*(19), 3737–3751. https://doi.org/10.1016/j.jmb.2016.07.010

Emma, P., Akre, R., Arthur, J., Bionta, R., Bostedt, C., Bozek, J., Brachmann, A., Bucksbaum, P., Coffee, R., Decker, F.-J., Ding, Y., Dowell, D., Edstrom, S., Fisher, A., Frisch, J., Gilevich, S., Hastings, J., Hays, G., Hering, Ph., … Galayda, J. (2010). First lasing and operation of an ångstrom-wavelength free-electron laser. *Nature Photonics*, *4*(9), 641–647. https://doi.org/10.1038/nphoton.2010.176

Emsley, P., & Cowtan, K. (2004). *Coot*: Model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography*, *60*(12), 2126–2132. https://doi.org/10.1107/S0907444904019158

Engilberge, S., Riobé, F., Di Pietro, S., Lassalle, L., Coquelle, N., Arnaud, C.-A., Pitrat, D., Mulatier, J.-C., Madern, D., Breyton, C., Maury, O., & Girard, E. (2017). Crystallophore: A versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence. *Chemical Science*, *8*(9), 5909–5917. https://doi.org/10.1039/C7SC00758B

Engilberge, S., Riobé, F., Wagner, T., Di Pietro, S., Breyton, C., Franzetti, B., Shima, S., Girard, E., Dumont, E., & Maury, O. (2018). Unveiling the Binding Modes of the Crystallophore, a Terbium-based Nucleating and Phasing Molecular Agent for Protein Crystallography. *Chemistry - A European Journal*, *24*(39), 9739–9746. https://doi.org/10.1002/chem.201802172

Engilberge, S., Wagner, T., Santoni, G., Breyton, C., Shima, S., Franzetti, B., Riobé, F., Maury, O., & Girard, E. (2019). Protein crystal structure determination with the crystallophore, a nucleating and phasing agent. *Journal of Applied Crystallography*, *52*(4), 722–731. https://doi.org/10.1107/S1600576719006381

Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, *357*(2), 289–298. https://doi.org/10.1016/j.ab.2006.07.027

Evans, P. (2005). Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography*, *62*(1), 72–82. https://doi.org/10.1107/S0907444905036693

Fischer, B., Perry, B., Phillips, G., Sumner, I., & Goodenough, P. (1993). Physiological consequence of expression of soluble and active hen egg white lysozyme in Escherichia coli. *Applied Microbiology and Biotechnology*, *39*(4–5), 537–540. https://doi.org/10.1007/BF00205047

García-Ruiz, J. M. (2003). Counterdiffusion Methods for Macromolecular Crystallization. In *Methods in Enzymology* (Vol. 368, pp. 130–154). Elsevier. https://doi.org/10.1016/S0076-6879(03)68008-0

García-Ruiz, J. M., & Moreno, A. (1994). Investigations on protein crystal growth by the gel acupuncture method. *Acta Crystallographica Section D Biological Crystallography*, *50*(4), 484–490. https://doi.org/10.1107/S0907444993014350

Giegé, R., Dock, A. C., Kern, D., Lorber, B., Thierry, J. C., & Moras, D. (1986). The role of purification in the crystallization of proteins and nucleic acids. *Journal of Crystal Growth*, *76*(3), 554–561. https://doi.org/10.1016/0022-0248(86)90172-7

Girard, É., Stelter, M., Vicat, J., & Kahn, R. (2003a). A new class of lanthanide complexes to obtain high-phasing-power heavy-atom derivatives for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography*, *59*(11), 1914–1922. https://doi.org/10.1107/S0907444903020511

Girard, É., Stelter, M., Vicat, J., & Kahn, R. (2003b). A new class of lanthanide complexes to obtain high-phasing-power heavy-atom derivatives for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography*, *59*(11), 1914–1922. https://doi.org/10.1107/S0907444903020511

Goldschmidt, L., Cooper, D. R., Derewenda, Z. S., & Eisenberg, D. (2007). Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Science*, *16*(8), 1569–1576. https://doi.org/10.1110/ps.072914007

Hampel, A., Labanauskas, M., Connors, P. G., Kirkegard, L., Rajbhandary, U. L., Sigler, P. B., & Bock, R. M. (1968). Single Crystals of Transfer RNA from Formylmethionine and Phenylalanine Transfer RNA's. *Science*, *162*(3860), 1384–1387. https://doi.org/10.1126/science.162.3860.1384

Hansen, S. B., Laursen, N. S., Andersen, G. R., & Andersen, K. R. (2017). Introducing site-specific cysteines into nanobodies for mercury labelling allows *de novo* phasing of their crystal structures. *Acta Crystallographica Section D Structural Biology*, *73*(10), 804–813. https://doi.org/10.1107/S2059798317013171

Hendrickson, W. A. (2014). Anomalous diffraction in crystallographic phase evaluation. *Quarterly Reviews of Biophysics*, *47*(1), 49–93. https://doi.org/10.1017/S0033583514000018

Hennig, M., Ruf, A., & Huber, W. (2011). Combining Biophysical Screening and X-Ray Crystallography for Fragment-Based Drug Discovery. In T. G. Davies & M. Hyvönen (Eds.), *Fragment-Based Drug Discovery and X-Ray Crystallography* (Vol. 317, pp. 115–143). Springer Berlin Heidelberg. https://doi.org/10.1007/128_2011_225

Hülsen, G., Broennimann, C., Eikenberry, E. F., & Wagner, A. (2006). Protein crystallography with a novel large-area pixel detector. *Journal of Applied Crystallography*, *39*(4), 550–557. https://doi.org/10.1107/S0021889806016591

Ingham, K. C. (1990). [23] Precipitation of proteins with polyethylene glycol. In *Methods in Enzymology* (Vol. 182, pp. 301–306). Elsevier. https://doi.org/10.1016/0076-6879(90)82025-W

Ireton, G. C., & Stoddard, B. L. (2004). Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallographica Section D: Biological Crystallography*, *60*(3), 601–605. https://doi.org/10.1107/S0907444903029664

Jiang, T., Roux, A., Engilberge, S., Alsalman, Z., Di Pietro, S., Franzetti, B., Riobé, F., Maury, O., & Girard, E. (2020). Tracking Crystallophore Nucleating Properties: Setting Up a Database for Statistical Analysis. *Crystal Growth & Design*, *20*(8), 5322–5329. https://doi.org/10.1021/acs.cgd.0c00556

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B.,

Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kabsch, W. (2010). *XDS*. *Acta Crystallographica Section D Biological Crystallography*, *66*(2), 125–132. https://doi.org/10.1107/S0907444909047337

Kahn, R., Fourme, R., Bosshard, R., Chiadmi, M., Risler, J. L., Dideberg, O., & Wery, J. P. (1985). Crystal structure study of *Opsanus tau* parvalbumin by multiwavelength anomalous diffraction. *FEBS Letters*, *179*(1), 133–137. https://doi.org/10.1016/0014-5793(85)80207-6

Kantardjieff, K. A., & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science*, *12*(9), 1865–1871. https://doi.org/10.1110/ps.0350503

Karplus, P. A., & Diederichs, K. (2012). Linking Crystallographic Model and Data Quality. *Science*, *336*(6084), 1030–1033. https://doi.org/10.1126/science.1218231

Kertis, F., Khurshid, S., Okman, O., Kysar, J. W., Govada, L., Chayen, N., & Erlebacher, J. (2012). Heterogeneous nucleation of protein crystals using nanoporous gold nucleants. *Journal of Materials Chemistry*, *22*(41), 21928. https://doi.org/10.1039/c2jm34527g

Khurshid, S., Saridakis, E., Govada, L., & Chayen, N. E. (2014). Porous nucleating agents for protein crystallization. *Nature Protocols*, *9*(7), 1621–1633. https://doi.org/10.1038/nprot.2014.109

Kolek, S. A., Bräuning, B., & Shaw Stewart, P. D. (2016). A novel microseeding method for the crystallization of membrane proteins in lipidic cubic phase. *Acta Crystallographica Section F Structural Biology Communications*, *72*(4), 307–312. https://doi.org/10.1107/S2053230X16004118

Larivière, L., Plaschka, C., Seizl, M., Wenzeck, L., Kurth, F., & Cramer, P. (2012). Structure of the Mediator head module. *Nature*, *492*(7429), 448–451. https://doi.org/10.1038/nature11670

Larson, S. B., Day, J. S., Cudney, R., & McPherson, A. (2007). A novel strategy for the crystallization of proteins: X-ray diffraction validation. *Acta Crystallographica Section D Biological Crystallography*, *63*(3), 310–318. https://doi.org/10.1107/S0907444906053303

Larsson, A., Jansson, A., Åberg, A., & Nordlund, P. (2011). Efficiency of hit generation and structural characterization in fragment-based ligand discovery. *Current Opinion in Chemical Biology*, *15*(4), 482–488. https://doi.org/10.1016/j.cbpa.2011.06.008

Lee, D. Y., Kim, K.-A., Yu, Y. G., & Kim, K.-S. (2004). Substitution of aspartic acid with glutamic acid increases the unfolding transition temperature of a protein. *Biochemical and Biophysical Research Communications*, *320*(3), 900–906. https://doi.org/10.1016/j.bbrc.2004.06.031

Legrandp, Soleilproxima1, Aishima, J., & CV-GPhL. (2019). *legrandp/xdsme: March 2019 version working with the latest XDS version (Jan 26, 2018)* (0.6.6). Zenodo. https://doi.org/10.5281/ZENODO.837885

Lye, R. C., Phillips, J. C., Kaplan, D., Doniach, S., & Hodgson, K. O. (1980). White lines in L-edge x-ray absorption spectra and their implications for anomalous diffraction studies of biological materials. *Proceedings of the National Academy of Sciences*, *77*(10), 5884–5888. https://doi.org/10.1073/pnas.77.10.5884

Maehashi, K., Matano, M., Irisawa, T., Uchino, M., Kashiwagi, Y., & Watanabe, T. (2012). Molecular characterization of goose- and chicken-type lysozymes in emu (Dromaius novaehollandiae): Evidence for extremely low lysozyme levels in emu egg white. *Gene*, *492*(1), 244–249. https://doi.org/10.1016/j.gene.2011.10.021

Mauracher, S. G., Molitor, C., Al-Oweini, R., Kortz, U., & Rompel, A. (2014). Latent and active *ab* PPO4 mushroom tyrosinase cocrystallized with hexatungstotellurate(VI) in a single crystal. *Acta Crystallographica Section D Biological Crystallography*, *70*(9), 2301–2315. https://doi.org/10.1107/S1399004714013777

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). *Phaser* crystallographic software. *Journal of Applied Crystallography*, *40*(4), 658–674. https://doi.org/10.1107/S0021889807021206

McGovern, R. E., Fernandes, H., Khan, A. R., Power, N. P., & Crowley, P. B. (2012). Protein camouflage in cytochrome c–calixarene complexes. *Nature Chemistry*, *4*(7), 527–533. https://doi.org/10.1038/nchem.1342

McGovern, R. E., McCarthy, A. A., & Crowley, P. B. (2014). Protein assembly mediated by sulfonatocalix[4]arene. *Chem. Commun.*, *50*(72), 10412–10415. https://doi.org/10.1039/C4CC04897K

McPherson, A. (1976). Crystallization of proteins from polyethylene glycol. *The Journal of Biological Chemistry*, *251*(20), 6300–6303.

McPherson, A., & Cudney, B. (2006a). Searching for silver bullets: An alternative strategy for crystallizing macromolecules. *Journal of Structural Biology*, *156*(3), 387–406. https://doi.org/10.1016/j.jsb.2006.09.006

McPherson, A., & Cudney, B. (2006b). Searching for silver bullets: An alternative strategy for crystallizing macromolecules. *Journal of Structural Biology*, *156*(3), 387–406. https://doi.org/10.1016/j.jsb.2006.09.006

McPherson, A., & Cudney, B. (2014). Optimization of crystallization conditions for biological macromolecules. *Acta Crystallographica Section F Structural Biology Communications*, *70*(11), 1445–1467. https://doi.org/10.1107/S2053230X14019670

McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica Section F Structural Biology Communications*, *70*(1), 2–20. https://doi.org/10.1107/S2053230X13033141

McPherson, A., & Shlichta, P. (1988). Heterogeneous and Epitaxial Nucleation of Protein Crystals on Mineral Surfaces. *Science*, *239*(4838), 385–387. https://doi.org/10.1126/science.239.4838.385

Mehrabi, P., Schulz, E. C., Dsouza, R., Müller-Werkmeister, H. M., Tellkamp, F., Miller, R. J. D., & Pai, E. F. (2019). Time-resolved crystallography reveals allosteric communication aligned with molecular breathing. *Science*, *365*(6458), 1167–1170. https://doi.org/10.1126/science.aaw9904

Molitor, C., Mauracher, S. G., & Rompel, A. (2015). Crystallization and preliminary crystallographic analysis of latent, active and recombinantly expressed aurone synthase, a polyphenol oxidase, from *Coreopsis grandiflora*. *Acta Crystallographica Section F Structural Biology Communications*, *71*(6), 746–751. https://doi.org/10.1107/S2053230X15007542

Nanev, C. N., Saridakis, E., Govada, L., Kassen, S. C., Solomon, H. V., & Chayen, N. E. (2019). Hydrophobic Interface-Assisted Protein Crystallization: Theory and Experiment. *ACS Applied Materials & Interfaces*, *11*(13), 12931–12940. https://doi.org/10.1021/acsami.8b20995

Ng, J. D., Gavira, J. A., & García-Ruíz, J. M. (2003). Protein crystallization by capillary counterdiffusion for applied crystallographic structure determination. *Journal of Structural Biology*, *142*(1), 218–231. https://doi.org/10.1016/S1047-8477(03)00052-2

Otálora, F., Gavira, J. A., Ng, J. D., & García-Ruiz, J. M. (2009). Counterdiffusion methods applied to protein crystallization. *Progress in Biophysics and Molecular Biology*, *101*(1–3), 26–37. https://doi.org/10.1016/j.pbiomolbio.2009.12.004

Parak, F., & Frauenfelder, H. (1993). Protein dynamics. *Physica A: Statistical Mechanics and Its Applications*, *201*(1–3), 332–345. https://doi.org/10.1016/0378-4371(93)90431-3

Pechkova, E., & Nicolini, C. (2002). Protein nucleation and crystallization by homologous protein thin film template. *Journal of Cellular Biochemistry*, *85*(2), 243–251. https://doi.org/10.1002/jcb.10123

Pompidor, G., D'Aléo, A., Vicat, J., Toupet, L., Giraud, N., Kahn, R., & Maury, O. (2008). Protein Crystallography through Supramolecular Interactions between a Lanthanide Complex and Arginine. *Angewandte Chemie International Edition*, *47*(18), 3388–3391. https://doi.org/10.1002/anie.200704683

Pompidor, G., Maury, O., Vicat, J., & Kahn, R. (2010). A dipicolinate lanthanide complex for solving protein structures using anomalous diffraction. *Acta Crystallographica Section D Biological Crystallography*, *66*(7), 762–769. https://doi.org/10.1107/S0907444910010954

Raynal, B., Lenormand, P., Baron, B., Hoos, S., & England, P. (2014). Quality assessment and optimization of purified protein samples: Why and how? *Microbial Cell Factories*, *13*(1), 180. https://doi.org/10.1186/s12934-014-0180-6

Reid, B. R., Koch, G. L. E., Boulanger, Y., Hartley, B. S., & Blow, D. M. (1973). Crystallization and preliminary X-ray diffraction studies on tyrosyl-transfer RNA synthetase from Bacillus stearothermophilus. *Journal of Molecular Biology*, *80*(1), 199–201. https://doi.org/10.1016/0022-2836(73)90241-6

Roux, A., Talon, R., Alsalman, Z., Engilberge, S., D'Aléo, A., Di Pietro, S., Robin, A., Bartocci, A., Pilet, G., Dumont, E., Wagner, T., Shima, S., Riobé, F., Girard, E., & Maury, O. (2021). Influence of Divalent Cations in the Protein Crystallization Process Assisted by Lanthanide-Based Additives. *Inorganic Chemistry*, *60*(20), 15208–15214. https://doi.org/10.1021/acs.inorgchem.1c01635

Salemme, F. R. (1972). A free interface diffusion technique for the crystallization of proteins for X-ray crystallography. *Archives of Biochemistry and Biophysics*, *151*(2), 533–539. https://doi.org/10.1016/0003-9861(72)90530-9

Santarsiero, B. D., Yegian, D. T., Lee, C. C., Spraggon, G., Gu, J., Scheibe, D., Uber, D. C., Cornell, E. W., Nordmeyer, R. A., Kolbe, W. F., Jin, J., Jones, A. L., Jaklevic, J. M., Schultz, P. G., & Stevens, R. C. (2002). An approach to rapid protein crystallization using nanodroplets. *Journal of Applied Crystallography*, *35*(2), 278–281. https://doi.org/10.1107/S0021889802001474

Saridakis, E., & Chayen, N. E. (2009). Towards a 'universal' nucleant for protein crystallization. *Trends in Biotechnology*, *27*(2), 99–106. https://doi.org/10.1016/j.tibtech.2008.10.008

Saridakis, E. E. G., Shaw Stewart, P. D., Lloyd, L. F., & Blow, D. M. (1994). Phase diagram and dilution experiments in the crystallization of carboxypeptidase G2. *Acta Crystallographica Section D Biological Crystallography*, *50*(3), 293–297. https://doi.org/10.1107/S0907444993013186

Saridakis, E., Khurshid, S., Govada, L., Phan, Q., Hawkins, D., Crichlow, G. V., Lolis, E., Reddy, S. M., & Chayen, N. E. (2011). Protein crystallization facilitated by molecularly imprinted polymers. *Proceedings of the National Academy of Sciences*, *108*(27), 11081–11086. https://doi.org/10.1073/pnas.1016539108

Sauter, C., Lorber, B., Kern, D., Cavarelli, J., Moras, D., & Giegé, R. (1999). Crystallogenesis studies on yeast aspartyl-tRNA synthetase: Use of phase diagram to improve crystal quality. *Acta Crystallographica Section D Biological Crystallography*, *55*(1), 149–156. https://doi.org/10.1107/S0907444998010890

Sauter, C., Lorber, B., McPherson, A., & Giege, R. (2012). Chapter 4.1. General methods. *International Tables for Crystallography (2012). Vol. F, Chapter 4.1, Pp. 99–121.. CRYSTALLIZATION*, 24.

Shinkai, Seiji., Araki, Koji., & Manabe, Osamu. (1988). NMR determination of association constants for calixarene complexes. Evidence for the formation of a 1:2 complex with calix[8]arene. *Journal of the American Chemical Society*, *110*(21), 7214–7215. https://doi.org/10.1021/ja00229a046

Stelter, M., Molina, R., Jeudy, S., Kahn, R., Abergel, C., & Hermoso, J. A. (2014). A complement to the modern crystallographer's toolbox: Caged gadolinium complexes with versatile binding modes. *Acta Crystallographica Section D Biological Crystallography*, *70*(6), 1506–1516. https://doi.org/10.1107/S1399004714005483

Stohrer, C., Horrell, S., Meier, S., Sans, M., von Stetten, D., Hough, M., Goldman, A., Monteiro, D. C. F., & Pearson, A. R. (2021). Homogeneous batch micro-crystallization of proteins from ammonium sulfate. *Acta Crystallographica Section D Structural Biology*, *77*(2), 194–204. https://doi.org/10.1107/S2059798320015454

Subramaniam, S. (2019). The cryo-EM revolution: Fueling the next phase. *IUCrJ*, *6*(1), 1–2. https://doi.org/10.1107/S2052252519000277

Szilágyi, A., & Závodszky, P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey. *Structure*, *8*(5), 493–504. https://doi.org/10.1016/S0969-2126(00)00133-7

Talon, R., Kahn, R., Durá, M. A., Maury, O., Vellieux, F. M. D., Franzetti, B., & Girard, E. (2011). Using lanthanoid complexes to phase large macromolecular assemblies. *Journal of Synchrotron Radiation*, *18*(1), 74–78. https://doi.org/10.1107/S0909049510036824

Talon, R., Nauton, L., Canet, J.-L., Kahn, R., Girard, E., & Gautier, A. (2012). Clicked europium dipicolinate complexes for protein X-ray structure determination. *Chemical Communications*, *48*(97), 11886. https://doi.org/10.1039/c2cc36982f

Taylor, G. L. (2010). Introduction to phasing. *Acta Crystallographica Section D Biological Crystallography*, *66*(4), 325–338. https://doi.org/10.1107/S0907444910006694

Thakur, A. S., Robin, G., Guncar, G., Saunders, N. F. W., Newman, J., Martin, J. L., & Kobe, B. (2007). Improved Success of Sparse Matrix Protein Crystallization Screening with Heterogeneous Nucleating Agents. *PLoS ONE*, *2*(10), e1091. https://doi.org/10.1371/journal.pone.0001091

Thorn, A., & Sheldrick, G. M. (2011). *ANODE*: Anomalous and heavy-atom density calculation. *Journal of Applied Crystallography*, *44*(6), 1285–1287. https://doi.org/10.1107/S0021889811041768

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., … Velankar, S. (2022). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, *50*(D1), D439–D444. https://doi.org/10.1093/nar/gkab1061

Vögeli, B., Engilberge, S., Girard, E., Riobé, F., Maury, O., Erb, T. J., Shima, S., & Wagner, T. (2018). Archaeal acetoacetyl-CoA thiolase/HMG-CoA synthase complex channels the intermediate via a fused CoA-binding site. *Proceedings of the National Academy of Sciences*, *115*(13), 3380–3385. https://doi.org/10.1073/pnas.1718649115

Volmer, M., & Weber, A. (1926). Keimbildung in übersättigten Gebilden. *Zeitschrift Für Physikalische Chemie*, *119U*(1), 277–301. https://doi.org/10.1515/zpch-1926-11927

Weaver, L. H., Grütter, M. G., Remington, S. J., Gray, T. M., Isaacs, N. W., & Matthews, B. W. (1984). Comparison of goose-type, chicken-type, and phage-type lysozymes illustrates the changes that occur in both amino acid sequence and three-dimensional structure during evolution. *Journal of Molecular Evolution*, *21*(2), 97–111. https://doi.org/10.1007/BF02100084

Wilson, J., Ristic, M., Kirkwood, J., Hargreaves, D., & Newman, J. (2020). Predicting the Effect of Chemical Factors on the pH of Crystallization Trials. *IScience*, *23*(6), 101219. https://doi.org/10.1016/j.isci.2020.101219

Young, L., Jernigan, R. L., & Covell, D. G. (1994). A role for surface hydrophobicity in protein-protein recognition: Using surface hydrophobicity to locate ligand attachment sites. *Protein Science*, *3*(5), 717–729. https://doi.org/10.1002/pro.5560030501

Zhou, R.-B., Cao, H.-L., Zhang, C.-Y., & Yin, D.-C. (2017). A review on recent advances for nucleants and nucleation in protein crystallization. *CrystEngComm*, *19*(8), 1143–1155. https://doi.org/10.1039/C6CE02562E

Résumé

Considérée comme l'une des méthodes les plus productives pour obtenir des informations structurales très précises sur les macromolécules biologiques, la cristallographie aux rayons X se heurte toujours à deux obstacles majeurs et ce malgré des avancées technologiques majeures: la production de cristaux de qualité/présentant une bonne diffraction et la résolution du problème de la phase inhérente à la méthode. C'est dans ce contexte que des complexes de lanthanides mono-cationiques ayant des propriétés de nucléation et permettant de résoudre le problème des phases, connus sous le nom de "Crystallophore" (Xo4), ont été développés.

Malgré les propriétés uniques qu'offre le TbXo4, les différentes études menées à son sujet montrent que son taux de réussite n'est pas de 100%, en particulier en ce qui concerne la cristallisation, d'où l'importance d'augmenter son taux de succès. L'objectif de cette thèse était de mieux comprendre les propriétés d'interaction et de nucléation du TbXo4 et son comportement à la surface des protéines.

Pour ce faire, nous avons abordé la question par une double approche. La première approche a consisté à explorer la chimie offerte par les complexes de lanthanide en évaluant les propriétés de nucléation de treize molécules dérivées du crystallophore, fournies par nos collaborateurs de l'ENS Lyon, sur un panel de trois protéines commerciales. Cette approche nous a permis de sélectionner une nouvelle molécule, TbXo4-OH, caractérisée par une extension propanol sur l'azote libre du macrocycle triazacyclononane, qui offre des propriétés de nucléation similaire à la molécule de référence TbXo4. Nous avons également développé un nouveau protocole d'utilisation des molécules.

La deuxième approche était une approche biologique qui a consisté à analyser les sites d'interaction entre TbXo4 et la cible protéique. Cette analyse a été combinée à de la mutagenèse (mutation ponctuelle), sur deux modèles de protéines, afin de confirmer les acides aminés ou groupes de résidus qui sont liés dans la liaison de la molécule. De plus, le présent travail a démontré la complémentarité entre la dynamique moléculaire et l'analyse de la structure cristalline.

Mots-clés: Cristallographie, diffraction rayons X, nucléation, complexes de lanthanide, crystallophore (Xo4), variants, mutagenèse, dynamique moléculaire.

Abstract

Considered as one of the most productive method for obtaining highly accurate structural information about biological macromolecules, X-ray crystallography still faces two major hurdles, despite tremendous technological advances that this method have encountered: production of high-quality/well diffracting crystals and solving the phase problem inherent to the method. Therefore, mono-cationic lanthanide complex with nucleating and phasing properties, known as “Crystallophore” (Xo4), has been developed.

Despite the important advantages that TbXo4 provides, the different studies undergone on it show that it is not 100% successful, in particular in term of crystallization, hence the importance of increasing its success. The objective of this thesis was to better understand the interaction and nucleation properties of TbXo4 and its behaviour on the protein surface.

To this end, we have addressed the question by a two-fold approach. The first approach has consisted in exploring the chemistry offered by lanthanide complexes by evaluating the nucleating properties of thirteen crystallophore variants, provided by our collaborators at ENS Lyon, on a panel of three commercial proteins. This approach has enabled us to select a new variant of crystallophore, TbXo4-OH, characterized by a propanol extension on the free nitrogen of the triazacyclononane macrocycle which provides similar nucleating properties compared to the reference molecule TbXo4. We have also developed a new protocol of use of these molecules.

The second approach was a biological approach that has consisted in analysing the interaction sites between TbXo4 and the protein target. This analysis was combined with mutagenesis (point mutation), on two protein models, in order to confirm amino acids or groups of residues that are related in the binding of the molecule.

Also, the present work has demonstrated the complementarity between molecular dynamics and crystal structure analysis.

Keywords: Crystallography, X-ray diffraction, nucleation, nucleating agents, lanthanide complex, crystallophore (Xo4), variants, mutagenesis, MD simulations.