

**Novel fusion tags for protein
production, purification and
research applications**

Mohamed Halawa

School of Pharmacy

University of Nottingham

December 2019

**Thesis for Doctorate of Philosophy degree to
the University of Nottingham**

Abstract

Soluble production of recombinant proteins is required for the study of structure and function of these proteins as well as for future therapeutic applications. Unfortunately, many proteins have a propensity to be expressed as insoluble proteins when expressed in a heterologous host, including *Escherichia coli*. The use of solubility tags is one of the strategies to overcome this difficulty.

The first aim of this thesis was to design and investigate new candidate fusion tags for protein expression in *E. coli* and as potential carrier mediated protein crystallization aids. Two fusion tag protein candidates had previously been identified from the Protein Data Bank (PDB) and literature searches based on criteria such as shape, size, ease of expression and crystallisation and the location of termini. These new fusion tags were the receiver domain of *Myxococcus xanthus* social motility protein *FrzS* (PDB code 2GKG) and the calponin homology domain from human β -spectrin (PDB code 1BKR). These tags proteins were fused to chosen human model proteins and peptides to assess expression, solubility and purification. Three groups of model proteins were selected for investigation: two weakly expressed proteins of unknown structures that display solubility issues (human Bardet Biedl protein 5, ubiquitin specific protease 20 zinc finger domain), two highly expressed proteins of known structures (ubiquitin carboxyl-terminal hydrolase isozyme L1, ubiquitin carboxyl-terminal hydrolase isozyme L3) and two peptides which are difficult to express without solubility tags, as the peptides have a high susceptibility for degradation. These two peptides (Ataxin residues 278-29 and α -synuclein residues 58-79) have been successfully obtained using Maltose binding protein (MBP) fusions. These proteins were expressed

as soluble proteins and were more amenable for purification when 2GKG and 1BKR tags were inserted at the fusion partners amino-terminal end. The results obtained in terms of solubility and purification with these novel tags compared favourably with results obtained using the widely used solubility tag Maltose binding protein (MBP) tag. Thus, these our two novel fusion tags have the ability to produce soluble proteins almost near to that level obtained by using MBP tag when they fused with these chosen model proteins and peptides. The fusion proteins produced were also investigated for protein crystallization, but no crystals have been obtained till now.

The second aim of this thesis was to design and investigate a new fusion tag utilizing the ubiquitin protein due to its high stability and solubility. Ubiquitin was engineered by insertion of a histidine tag into the loop between strands 3 and 4 to be used as affinity and solubility tag. The ubiquitin modified with a histidine tag inserted between the strands 3 and 4 (UBHIS3S4) was expressed and purified alone in a high yield but when fused to two model proteins BBS5(1-270) and USP20 ZNFUBP (1-108) showed relatively a low yield. These results suggested that UBHIS3S4 protein was not favourable for use as affinity and solubility tag in these two instances.

In addition, the ubiquitin protein was designed and evaluated as protein scaffold due to its high stability and solubility. A peptide sequence from a constrained peptide library had previously been identified by next generation phage display technology that interacts with the N-terminal domain of the ubiquitin specific protease 11 (USP11) residues (1-244). This peptide sequence was inserted between strands 3 and 4 of ubiquitin to assess the suitability of ubiquitin as a scaffold for peptide display. The modified ubiquitin was expressed and purified and binding studies with USP11 (1-244)

were performed by isothermal titration calorimetry which confirmed the binding. The results suggested that ubiquitin is suitable as a scaffold to display peptides which are at least 9 amino acids length between strands 3 and 4. Use of ubiquitin as a protein scaffold could serve as the basis for the production of peptidomimetic drugs and for different studies on drug discovery as anticancer drugs.

Acknowledgement

I would like to take this opportunity to first and foremost thank my supervisors Dr. Ingrid Dreveny and Prof. Jonas Emsley for the continuous support of my research, for their patience, motivation, enthusiasm and immense knowledge. I want to take this chance also to thank my parents and my wife for their unconditional support throughout my degree. I wish to thank my family, my colleagues and all my friends for providing a loving environment for me.

List of contents

1.	CHAPTER 1: INTRODUCTION	1
1.1	CHALLENGES AND FACTORS AFFECTING THE PRODUCTION OF RECOMBINANT PROTEIN	3
1.2	CHARACTERIZATION OF RECOMBINANT PROTEIN PRODUCTION	6
1.3	STRATEGIES TO CONTROL STABILITY OF THE RECOMBINANT PROTEIN PRODUCTION	6
1.3.1	Bioinformatical assessment and control of protein aggregation	6
1.3.2	Modification of genes of interest by site directed mutagenesis to control aggregation ...	7
1.3.3	Expression systems	7
1.3.4	Alterations in expression conditions.....	8
1.3.5	Using fusion tags.....	8
1.4	BACKGROUND TO THIS STUDY	15
1.4.1	Novel potential solubility of new suggested tags (2GKG and 1BKR).....	15
1.4.2	Chosen model proteins used for investigation with the novel tags.....	20
1.4.3	Design and investigate the suitability of the ubiquitin protein for protein affinity, solubility and protein scaffolds	26
1.5	THE AIMS OF THIS STUDY	30
1.6	Specific aims	31
2.	CHAPTER 2: MATERIALS AND METHODS.....	34
2.1	MATERIALS	34
2.1.1	Chemicals	34
2.1.2	Enzymes.....	34

2.1.3	Bacterial Strains.....	34
2.1.4	Plasmids.....	34
2.1.5	DNA ladders	37
2.1.6	Media.....	37
2.1.7	Media for auto induction	37
2.1.8	Cell strains	37
2.1.9	Preparation of agarose gel	38
2.1.10	Miniprepation Buffers plasmid for DNA isolation.....	38
2.1.11	Buffers for protein purification.....	38
2.1.12	Crystallisation solutions	39
2.2	METHODS	40
2.2.1	Bacteria strains growth.....	40
2.2.2	Transformation of competent <i>E coli</i> strains	40
2.2.3	Measurement of DNA concentration	40
2.2.4	DNA plasmid isolation (mini-preparation):.....	40
2.2.5	Molecular cloning methods.....	41
2.2.6	Site directed mutagenesis	46
2.2.7	Transformants analysis (Mini digestion test)	47
2.2.8	Gene sub-cloning	47
2.2.9	Protein production.....	47
2.2.10	Protein purification	49

2.2.11	Measurement of protein concentration.....	51
2.2.12	Protein characterization methods	51
2.2.13	Other methods.....	56
2.2.14	Cryo-cooling	62
2.2.15	Bioinformatics servers	62
3.	CHAPTER 3 RESULTS AND DISCUSSION OF PREPARATION AND INVESTIGATION OF NOVEL CANDIDATE TAGS 1BKR AND 2GKG	64
3.1	Cloning of 1BKR and 2GKG.....	65
3.2	Expression, solubility and purification test of 1BKR and 2GKG tags in small-scale auto induction	65
3.3	Expression, solubility and purification of 1BKR and 2GKG tags in large-scale cultures.....	68
3.4	Site directed mutagenesis of 1BKR tag	71
3.5	Crystallization trials of 1BKRwildtype,1BKRL17D, 1BKRL17S and 2GKG for crystallization	75
3.6	DISCUSSION OF 1BKR AND 2GKG PROTEIN EXPRESSION AND CRYSTALLISATION RESULTS.....	76
4.	CHAPTER 4 RESULTS AND DISCUSSION OF THE FUSION OF 1BKR AND 2GKG TAGS TO MODEL PROTEINS UCHL1 AND UCHL3.....	79
4.1	RESULTS OF INVESTIGATION OF UCHL1-TAG FUSIONS	81
4.1.1	Cloning of 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1 constructs	81
4.1.2	Expression, solubility and small-scale purification of 1BKR-UCHL1, 2GKG-UCHL1	82
4.1.3	Large-scale expression and purification of UCHL1 protein- solubility tag fusions	84
4.1.4	Investigation of UCHL1 protein- solubility tag fusions using small Angle X ray.....	90
4.1.5	Crystallisation trials of UCHL1 protein- solubility tag fusions.....	94
4.2	RESULTS OF THE INVESTIGATION OF UCHL3-TAG FUSIONS	96

4.2.1	Cloning of 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3 constructs	96
4.2.2	Expression, solubility and small-scale purification of 1BKR-UCHL3, 2GKG-UCHL3	96
4.2.3	Large-scale expression and purification of UCHL3 protein- solubility tag fusions	98
4.2.4	Analysis of UCHL3 protein- solubility tag fusions using small Angle X ray scattering	104
4.2.5	Crystallisation trials of UCHL3 protein-solubility tag fusions	108
4.3	Discussion of the results of UCHL1 and UCHL3 proteins with the tags.....	110
5.	CHAPTER 5 RESULTS AND DISCUSSION OF EFFECT OF NOVEL TAGS ON THE EXPRESSION AND PURIFICATION OF ATAXIN-3 (278-291) AND ALPHA SYNUCLEIN (58-79) PEPTIDES	115
5.1	RESULTS AND INVESTIGATION OF ATAXIN-3 (278-291) PEPTIDE -TAG FUSIONS	116
5.1.1	Cloning of 1BKR-Ataxin (278-291), 2GKG-Ataxin (278-291)	116
5.1.2	Expression, solubility and small-scale purification of 1BKR-Ataxin (278-291), 2GKG-Ataxin (278-291) and MBP-Ataxin (278-291).....	117
5.1.3	Large-scale expression and purification of Ataxin (278-291) peptide-solubility tag.....	119
5.1.4	Crystallisation trials of Ataxin (278-291) peptide- solubility tag fusions	125
5.2	RESULTS AND INVESTIGATION OF A-SYNUCLEIN (58-79) PEPTIDE-TAG FUSIONS.....	128
5.2.1	Cloning of 1BKR-alpha synuclein (58-79), 2GKG-alpha synuclein (58-79) and MBP-alpha synuclein (58-79) constructs	128
5.2.2	Expression, solubility and small-scale purification of 1BKR-alpha synuclein, 2GKG-alpha synuclein and MBP-alpha synuclein.....	129
5.2.3	Large-scale expression and purification of alpha synuclein (58-79) peptide-solubility ..	131
5.2.4	Crystallization trials of alpha synuclein (58-79) peptide- solubility tag fusions	137
5.3	DISCUSSION OF THE RESULTS OF ATAXIN (278-291) AND ALPHA SYNUCLEIN (58-79)	139

6.	CHAPTER 6: RESULTS AND DISCUSSION OF NOVEL TAGS FUSED TO PROTEINS OF UNKNOWN STRUCTURE, USP20 ZNFUBP (1-108) AND BBS5 (1-270)	142
6.1	RESULTS OF INVESTIGATION OF THE FIRST MODEL PROTEIN USP20 ZNFUBP (1-108)	143
6.1.1	Cloning constructs of 1BKR-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and MBP-USP20ZNFUBP (1-108).....	143
6.1.2	Expression, solubility and small-scale purification of 1BKR-USP20ZNFUBP (1-108), 2GKG-USP20 ZNFUBP (1-108) and His-USP20ZNFUBP (1-108).....	144
6.1.3	Large-scale expression and purification of USP20ZNFUBP (1-108) protein tag fusions .	146
6.1.4	Crystallisation trials of USP20ZNFUBP(1-108) protein- solubility tag fusions	152
	Table 24. Results of crystallisation using different crystal screens for 5 mg/mL 1BKR-USP20 ZNFUBP (1-108) protein after one week.	153
6.2	INVESTIGATION OF THE BBS5 (1-270) PROTEIN.....	155
6.2.1	Cloning of 1BKR-BBS5 (1-270), 2GKG BBS5 (1-270) and MBP- BBS5 (1-270)	155
6.2.2	Expression, solubility and small scale purification of 1BKR-BBS5 (1-270), 2GKG-BBS5 (1-270) and His-BBS5 (1-270) protein	156
6.2.3	Large-scale expression and purification of BBS5 (1-270) protein-solubility tag fusions .	158
6.2.4	Crystallisation trial BBS5 (1-270) protein- solubility tag fusions	164
6.3	DISCUSSION OF THE RESULTS OF USP20ZNFUBP (1-108) AND BBS5 (1-270) FUSIONS	165
7.	CHAPTER 7: RESULTS AND DISCUSSION OF UBIQUITIN PROTEIN APPLICATIONS AS AFFINITY, SOLUBILITY TAG OR PROTEIN SCAFFOLD	169
7.1	UBIQUITIN PROTEIN AS AFFINITY AND SOLUBILITY TAG	170
7.1.1	Cloning of a UbS3S4His tag construct.....	170
7.1.2	UbS3S4His expression and purification	171
7.1.3	Assessing UbS3S4His protein thermal stability by Thermoflour assay	173

7.1.4	Crystallization trials of UbS3S4His	174
7.2	INVESTIGATION OF UbS3S4His-USP20ZNFUBP (1-108) AND UbS3S4His -BBS5 (1-270).....	175
7.2.1	Cloning of UbS3S4His-USP20ZNFUBP (1-108) and UbS3S4His-BBS5 (1-270).....	175
7.2.2	Large-scale expression and purification of UbS3S4His-USP20ZNFUBP (1-108)	175
7.2.3	Large-scale expression and purification of UbS3S4His-BBS5 (1-270)	177
7.3	THE USE OF THE UBIQUITIN PROTEIN AS A PROTEIN SCAFFOLD	179
7.3.1	Cloning of UbpeptideS3S4 and USP11 N-terminal domain (1-244).....	179
7.3.2	Expression and Purification of Ubiquitin peptide S3S4	180
7.3.3	Expression and purification of USP11 domain residues (1-244)	182
7.3.4	Interaction studies of USP11 (1-244) with UbpepS3S4 by isothermal titration	184
7.4	DISCUSSION OF RESULTS OF UBIQUITIN MODIFICATIONS.....	186
8.	CHAPTER 8: CONCLUSION	188
9.	CHAPTER 9: FUTURE OUTLOOK	194
10.	REFERENCES	197

List of the Figures

Figure 1. Flowchart showing the process of recombinant protein production started by cloning then expression and purification. Figure adapted from (13).....	2
Figure 2. Graphical representation showing the protein folding. During folding, hydrophobic amino acids (represented as green spheres) form the protein core while hydrophilic amino acids (represented as pink spheres) form water exposed surfaces of the protein Figure adapted from (14).....	3
Figure 3. Diagram represents misfolding of native protein under certain circumstances such as pH and temperature change. Figure adapted from (27)	5
Figure 4 Schematic of a plasmid encoding a target protein and fusion tag.	9
Figure 5. Possible mechanism of solubility tag as micelle. The solubility tag (indicated in yellow) and target protein (indicated in red). During this mechanism, separation of the target protein from the solvent occurs while the domains of the soluble protein simultaneously face outward by micelle-like structures.....	10
Figure 6. Graphical representation showing that a solubility tag displays intrinsic chaperone activity to protect its fusion protein from aggregation. Figure adopted from (46).....	11
Figure 7. Distribution of 1BKR and 2GKG surface charges. The negative charges are in red colour, the positive charges are in blue colour and the white regions are neutral. (made in Pymol software).	19
Figure 8. Cartoon diagram of the crystal structure of two candidate tags A) 1BKR, B) 2GKG (made in Pymol software).....	19
Figure 9. Cartoon diagram of the crystal structure of UCHL1 with the PDB code: 2etl (made using Pymol software). The N terminus represented in green colour while the C terminus represented in yellow colour.	21
Figure 10. Cartoon diagram of the crystal structure of UCHL3 with the PDB code: 1UCH (made using Pymol software). The N terminus represented in red colour while the C terminus represented in yellow colour.	21
Figure 11. Cartoon diagram of the crystal structure of MBP-ataxin (278-291) PDB code: 4YS9 (made using pymol software). The MBP tag represented in red colour while ataxin (278-291) represented in green colour.	23
Figure 12. Cartoon diagram of the crystal structure of MBP- α - synuclein (58-79) PDB code: 3Q28 (made using Pymol software). The MBP tag represented in red colour while α -synuclein peptide (58-79) represented in violet colour.....	24
Figure 13. Cartoon diagram of the crystal structure of ubiquitin protein (PDB code 1UBI). The figure represents the loop between strands 3 and strand 4 in yellow colour (made using Pymol software).	28
Figure 14. Example of the primer design of Annealing Oligonucleotides Primers.	44

Figure 15. Thermogram revealing different protein states including native, denatured and aggregated states. Figure adapted from (147)	52
Figure 16.Example of data analysis of the Guinier Plot. Figure adapted from (153).	54
Figure 17. Example of Dimensionless Kratky plots. Figure adapted from (154).	54
Figure 18.Example of proteins profiles with different structural features in pair distribution function. Figure adapted from (160).	55
Figure 19. Diagrammatic representative of isothermal titration calorimetry. Figure adapted from (170).	57
Figure 20.Schematic representation of Vapor diffusion crystallization technique. Figure adapted from (183)	59
Figure 21. Diagrammatic description of a two-dimensional phase diagram showing the two states of protein solution which are under saturation and super saturation states. Figure adapted from (188).....	60
Figure 22. A: Analysis of the un-induced and induced fractions from protein expression experiment of 1BKR tag on SDS-PAGE gel. B: un-induced and induced samples of 2GKG protein were run on a 15 % SDS-PAGE gel C: SDS-PAGE gel of the 1BKR tag solubility test. D: The SDS-PAGE gel of the 2GKG tag solubility revealed protein bands in both the insoluble and soluble fractions E: The SDS-PAGE gel of 1BKR tag purification indicated that the 1BKR tag had been successfully purified. F: The SDS-PAGE gel of 2GKG tag purification revealed a band corresponding to that expected for 2GKG tag.	67
Figure 23. A: Chromatogram of the gel filtration purification of 2GKG. The Chromatogram showed a monodisperse peak containing the protein target. B: SDS-PAGE gel investigation of the peak fractions revealed bands, which corresponded to the expected size of the 2GKG protein (14 kDa). C: Chromatogram of the 1BKR protein after gel filtration purification. D: SDS-PAGE gels showed an intense band below the marker band of 14.3 kDa which corresponds to the 1BKR protein.	69
Figure 24. Standard curve for Superdex 75(16 /60) column showing the elution volumes of the different molecular weight of proteins used as standards.	70
Figure 25. Cartoon diagram of the crystal structure of 1BKR , the leucine residue number 17 highlighted in yellow colour (made in Pymol software).	70
Figure 26. A: Gel filtration chromatogram of 1BKRL17S revealed one peak at 90 ml as elution volume corresponding to 13 kDa based on the standard curve. B: SDS-PAGE gel analysis revealed that the protein was approximately 13 kDa that corresponded to 1BKRL17S C: Gel filtration chromatogram of 1BKRL17D showed a high peak containing the target protein. D: SDS-PAGE gel result showed bands that corresponded to 1BKRL17D (13 kDa).	72
Figure 27. Sample of raw data from the ThermoFlour assays showing the measurement of the T _m which is the temperature where the protein unfolds revealing its hydrophobic core which is defined as the midpoint of the curve. The melting temperature represented by the turquoise colour while the blue colour after turquoise colour refers to protein aggregation. A: the first measurement B: second measurement, C: third measurement.	74

Figure 28. Bar graph representing the melting temperature measurements from the ThermoFluor experiments of 1BKR wild type, 1BKRL17D, 1BKRL17S and 2GKG proteins.	74
Figure 29. (A) Crystal of 1BKR wild protein in condition: conc. Salt: 1.5M sodium citrate tribasic dehydrate, conc. Buffer: 0.01 M sodium borate, pH 8.5. (B) Crystals of IBKR wild protein in condition: conc. Salt: 1.2M sodium citrate tribasic dehydrate, conc. Buffer: 0.01 M sodium borate, pH 8.5. (C) Crystal from 2GKG protein on JCSG+ screen on the condition is 0.02 M Calcium chloride dehydrate, 0,1 M sodium acetate, 30% v/v MPD pH 4.6	75
Figure 30.A: The SDS-PAGE gel analysis of expression levels of 1BKR-UCHL1, 2GKG- UCHL1 and MBP-UCHL1. SDS-PAGE indicated all the proteins were expressed as indicated by the presence of bands after-induction where no bands were present before induction. B: SDS-PAGE gel analysis of the soluble fractions of 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1. C: SDS-PAGE analysis of samples after purification of 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1	83
Figure 31. A: Affinity chromatography of 1BKR-UCHL1. B: SDS-PAGE gel from Ni column showed 1BKR UCHL1 has 39.6 kDa. C: Chromatogram of 1BKR-UCHL1 after gel Filtration purification. D: SDS-PAGE gels showed a band at around 40 kDa, which is analogous to that of 1BKR UCHL1.	85
Figure 32. A: Chromatogram of the affinity chromatography showed a broad peak which contains 2GKG-UCHL1 protein. B: SDS-PAGE gel from Ni column showed 2GKG-UCHL1. C: Gel filtration chromatogram of 2GKG-UCHL1. D: Analysis of the peak fractions on SDS-PAGE gels revealed a band of around 40 kDa, this band was the correct predicted size for 2GKG-UCHL1.	87
Figure 33.A.The chromatogram of affinity chromatography of MBP-UCHL1 Chromatogram showed a prominent peak which contained the target protein. B: Analysis of the peak fractions and insoluble pellets on SDS-PAGE gel revealed protein bands which corresponded to the size of MBP-UCHL1 of approximately 66.2 kDa. C: The chromatogram of gel filtration purification of MBP-UCHL1. D: SDS-PAGE gel result showed bands that corresponded to MBP-UCHL1 (66.2 kDa).	89
Figure 34.:A: The Guinier plot of the 1BKR UCHL1 protein. B: Guinier plot of the 2GKG UCHL1 protein. C: Guinier plot of MBP-UCHL1 protein.	91
Figure 35.A: Dimensionless Kratky plot of 1BKR-UCHL1 protein, B: Dimensionless Kratky plot of 2GKG-UCHL1 protein. C: Dimensionless Kratky plot of MBP-UCHL1 protein.	92
Figure 36.A: Pair-Distribution Function P(r) plot of 1BKR-UCHL1 protein, B: Pair-Distribution Function P(r) plot of 2GKG-UCHL1 protein. C: Pair-Distribution Function P(r) plot of MBP-UCHL1 protein.	92
Figure 37. A: The prediction of the maximum linear dimension (Dmax) by pymol software of models of the 1BKR-UCHL1 proteins from their crystal structures which is equal 75 .7 Å. B: The prediction of the maximum linear dimension (Dmax) by pymol software of the 2GKG-UCHL1 protein from their crystal structures which was equal to 79.2 Å. C: The prediction of the maximum linear dimension (Dmax) by pymol software of the MBP-UCHL1 protein from their crystal structures which was equal to 86 Å.	93

Figure 38.A: SDS-PAGE gel of expression test of the three proteins 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3. B: SDS-PAGE gel of solubility test of the three proteins 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3. C: SDS-PAGE gel of purification test of the three proteins 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3. 97

Figure 39.A:Affinity chromatography and SDS-PAGE gel of 1BKR-UCHL3.B: SDS-PAGE gel from Ni column. C: Chromatogram of 1BKR-UCHL3 gel Filtration purification. D: SDS-PAGE gels represented a band at around 40 kDa which is a 1BKR UCHL3.....99

Figure 40. A. The affinity column chromatogram. B: SDS-PAGE gel from a Ni column run revealed a band at 40 kDa analogous to the predicted size of 2GKG-UCHL3 with some contaminating proteins. C: Chromatogram of 2GKG-UCHL3 gel filtration purification. D: SDS-PAGE gels revealed a band at around 40 kDa that is 2GKG-UCHL3. 101

Figure 41.A: The affinity chromatography chromatogram of MBP-UCHL3. B: SDS-PAGE gel from Ni column showed MBP-UCHL1 at approximately 67 kDa C: Chromatogram of MBP-UCHL3 gel filtration purification. D: SDS-PAGE gels represented a band at about 67 kDa, this band is the MBP-UCHL3 protein.....103

Figure 42.A: The Guinier plot of 1BKR-UCHL3 protein. B: The Guinier plot of 2GKG-UCHL3 protein .C The Guinier plot of MBP-UCHL3 protein. 105

Figure 43.A: Dimensionless Kratky plot of 1BKR-UCHL3 protein. B: Dimensionless Kratky plot 2GKG-UCHL3 protein. C: Dimensionless Kratky plot of MBP-UCHL3 protein. 106

Figure 44.A: Pair-Distribution Function P(r) plot of 1BKR-UCHL3 protein. B: Pair-Distribution Function P(r) plot of 2GKG-UCHL3 protein. C: Pair-Distribution Function P(r) plot of MBP-UCHL3.....106

Figure 45.The prediction of the maximum linear dimension (Dmax) by Pymol software of the 1BKR-UCHL3 protein from their crystal structures, which is equal to 70.9 Å. B: The prediction of the maximum linear dimension (Dmax) by pymol software of the 2GKG and UCHL3 proteins from their crystal structures which was equal to 77 Å. C: The prediction of the maximum linear dimension (Dmax) by pymol software of the MBP and UCHL3 proteins from their crystal structures which is equal 83.6 Å.....107

Figure 47.The crystal from 2GKG -UCHL3 but no diffraction was obtained. The condition was 0.2 M Calcium chloride dihydrate 0.1 M Tris 20 % w/v PEG 6000, pH=8.....109

Figure 48. Graph showing a relative comparison of the yields of 1BKR-UCHL1, 2GKG- UCHL1 and MBP-UCHL1. (no error bar because it just only one experiment). 113

Figure 49.Graph showed the comparison of yields of 1BKR-UCHL3, 2GKG- UCHL3 and MBP-UCHL3 (no error bar because it just only one experiment).....114

Figure 50.A: The SDS-PAGE gel of expression test indicated all the proteins were expressed and have extra band in the post-induction fractions. B: The SDS-PAGE gel of solubility test of MBP-Ataxin (278-291), 2GKG-Ataxin (278-291) and 1BKR-Ataxin (278-291). C: The SDS-PAGE gel of purification test of the three proteins MBP-Ataxin (278-291), 2GKG- Ataxin (278-291) and 1BKR-Ataxin (278-291).....118

Figure 51.A: Nickel purification chromatogram of 1BKR-Ataxin (278-291) protein. B: SDS-PAGE gel from nickel column purification showed the pure protein. C: Gel filtration purification chromatogram of 1BKR-Ataxin (278-291) protein. D: SDS-PAGE gel showed purer band compared to Ni column purification with the correct size of molecular weight refer to protein marker..... 120

Figure 52.A: The nickel column purification chromatogram of 2GKG-Ataxin (278-291) protein. B: the purified fractions analysis by SDS-PAGE gel revealed the existence of a purified protein. C: Chromatogram after gel filtration purification of 2GKG-Ataxin (278-291) protein. D: SDS-PAGE gel of 2GKG-Ataxin (278-291) protein after gel filtration purification..... 122

Figure 53.A: The nickel column chromatogram of MBP-Ataxin (278-291) protein. B: SDS-PAGE gel from nickel column purification of MBP-Ataxin (278-291) protein. C: Gel filtration purification chromatogram of MBP-Ataxin (278-291) protein. D: SDS-PAGE gel of MBP-Ataxin (278-291) protein after gel filtration purification..... 124

Figure 54.Potential crystals of the 1BKR-Ataxin (278-291) protein. The condition was PEG 400 40% Citric acid Buffer, pH 6.5. 125

Figure 55. The crystal-like objects from 2GKG-Ataxin (278-291) protein. 126

Figure 56.A:The SDS-PAGE gel of expression test of the three proteins 1BKR-alpha synuclein (58-79), 2GKG-alpha synuclein (58-79) and MBP with alpha synuclein (58-79). B: The SDS-PAGE gel of solubility test of the three proteins 1BKR- alpha synuclein (58-79), 2GKG- alpha synuclein (58-79) and MBP with alpha synuclein (58-79). C: The SDS-PAGE gel of purification test of the three proteins MBP- alpha synuclein (58-79) 2GKG-alpha synuclein (58-79) and 1BKR-alpha synuclein (58-79). 130

Figure 57. A: Nickel column chromatogram of 1BKR-alpha synuclein (58-79). B: The SDS-PAGE gel of purified 1BKR-alpha synuclein (58-79) samples. C: Gel filtration purification chromatogram of 1BKR-alpha synuclein (58-79). D: 1BKR-alpha synuclein (58-79) was expected on 15 kDa as shown on the SDS-PAGE gel. 132

Figure 58. A: Nickel affinity purification chromatogram of 2GKG-alpha synuclein (58-79). B: The SDS-PAGE gel indicates the expression and purification of the 2GKG alpha synuclein (58-79) of the predict size 15 .8 kDa. C: Gel filtration purification chromatogram of 2GKG-alpha synuclein. D: 2GKG alpha synuclein (58-79)) was expected on 15.8 kDa as shown on the SDS-PAGE gel..... 134

Figure 59.A: The chromatogram of affinity chromatography results of MBP-alpha synuclein (58-79). B: SDS-PAGE gel result C: Nickel column purification Chromatogram of MBP-alpha synuclein (58-79). D: SDS-PAGE gel result showed bands that correspondent to MBP-alpha synuclein (58-79) (42 kDa). 136

Figure 60. Precipitate from 1BKR-alpha synuclein (58-79) protein. 137

Figure 61.The crystal from 2GKG-alpha synuclein (58-79) protein in the PACT screen (0.2 M Sodium chloride 0.1 M MES,20 % w/v PEG 6000,pH- 6.0). 138

Figure 62. Graph showing a comparison of the yield of 1BKR- Ataxin (278-291), 2GKG- Ataxin (278-291) and MBP Ataxin (278-291) constructs after gel filtration purification (no error bar because it just only one experiment). ... 141

Figure 63. Graph showed the comparison of yield of MBP- alpha synuclein (58-79) 2GKG- alpha synuclein (58-79) and 1BKR alpha synuclein (58-79) (no error bar because it just only one experiment).....	141
Figure 64.A: The SDS-PAGE gel of expression test of the three proteins His-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and 1BKR-USP20ZNFUBP (1-108). B: The SDS-PAGE gel of solubility test of the three proteins His-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and 1BKR USP20ZNFUBP (1-108). C: The SDS-PAGE gel of purification test of the three proteins His-USP20 ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and 1BKR-USP20 ZNFUBP (1-108).	145
Figure 65.A: The chromatogram affinity chromatography of 1BKR-USP20ZNFUBP (1-108). B: SDS-PAGE gel from Ni column showed 1BKR-USP20 ZNFUBP (1-108) C: Chromatogram of 1BKR-USP20 ZNFUBP (1-108) gel Filtration Purification. D: SDS-PAGE gels revealed a band at around 25 kDa that is a 1BKR-USP20 ZNFUBP (1-108).....	147
Figure 66.A: Chromatogram of affinity chromatography of 2GKG USP20 ZNFUBP (1-108). B: SDS-PAGE gel from Ni column showed the two peaks contained the 2GKG-USP20ZNFUBP (1-108) protein C: first gel filtration purification chromatogram of 2GKG-USP20ZNFUBP (1-108) from first peak of nickel purification. D: Second gel filtration purification chromatogram of 2GKG-USP20ZNFUBP (1-108) from second peak of nickel purification. E: SDS-PAGE gels showed the two gel filtration purifications only the fractions second peak were collected for crystallization trials.....	149
Figure 67. A: Nickel column chromatogram showed a single peak. B: SDS-PAGE gel showing the MBP-USP20ZNFUBP protein was expressed. C: Gel filtration chromatogram of MBP-USP20ZNFUBP (1-108). D: The SDS-PAGE gel of MBPUSP20ZNFUBP showed the single protein was purified.	151
Figure 68. Heavy precipitate in JCSG plus screen from 1BKR-USP20 ZNFUBP protein. The condition was 0.2 M Zinc acetate dihydrate 0.1 M Sodium acetate, 10 % w/v PEG 3000, pH 4.5.	152
Figure 69. Crystals from the 2GKG-USP20 ZNFUBP (1-108) protein on Morpheus screen condition: 0.12 M Ethylene glycols 0.1 M Buffer System 1, 50 % v/v Precipitant Mix 1, pH 6.5.....	153
Figure 70.A: The SDS-PAGE gel of expression test of the three proteins His-BBS5 (1-270), 2GKG-BBS5 (1-270) and 1BKR-BBS5 (1-270). B: The SDS-PAGE gel of solubility test of the three proteins His-BBS5 (1-270) 2GKG-BBS5 (1-270) and 1BKR BBS5 (1-270). C: The SDS-PAGE gel of purification test of the three proteins His- BBS5 (1-270), 2GKG- BBS5 (1-270) and 1BKR BBS5 (1-270).	157
Figure 71.A: Nickel column chromatogram of 1BKR-BBS5 (1-270). B: The SDS-PAGE gel indicates contaminate with 1BKR-BBS5 (1-270) protein.. C: The chromatogram of gel filtration showed an elution volume of protein at 86 mL with a small peak after the void volume. D: Th SDS-PAGE gel after gel filtration from 1BKR-BBS5 (1-270).	159
Figure 72A: Nickel column chromatogram of 2GKG-BBS5 (1-270). B: The SDS-PAGE gel after Nickel column purification of 2GKG-BBS5 (1-270) protein. C: The gel filtration chromatogram of 2GKG-BBS5 (1-270) protein. D: The SDS-PAGE gel after gel filtration purification of 2GKG-BBS5 (1-270) protein.	161

Figure 73. A: The affinity chromatography chromatogram of MBP-BBS5 (1-270). B: SDS-PAGE after Nickel column purification of MBP-BBS5 (1-270). C: The chromatogram of gel filtration purification of MBP-BBS5 (1-270). D: SDS-PAGE gel result showing bands that correspondent to MBP-BBS5 (1-270) (71 kDa).	163
Figure 74. Comparison of yields of 1BKR-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP(1-108) and MBP-USP20ZNFUBP (1-108) constructs after gel filtration purification (no error bar because it just only one experiment).	167
Figure 75. Graph showed the comparison of yields of 1BKR- BBS5 (1-270), 2GKG- BBS5 (1-270), MBP- BBS5 (1-270) constructs after gel filtration purification (no error bar because it just only one experiment).	168
Figure 76 . Cartoon diagram of the ubiquitin protein structure (PDB code 1UBI) and the yellow colour represents the loop between strands 3 and 4 where the histidine tag was inserted.....	171
Figure 77. A: The Nickel column purification chromatogram of UbS3S4His. B: SDS-PAGE gel result showed bands that corresponded to UbS3S4His at about 8.5 kDa. C: The gel filtration chromatogram of UbS3S4His. D: SDS-PAGE gel result showed bands that correspondent to UbS3S4His at about 8.5 kDa.	172
Figure 78. Sample of raw data from the Thermofluor assay. The graph of the fluorescence against temperature of UbS3S4His	173
Figure 79. A: SDS-PAGE after the affinity chromatography. B: The chromatogram of the gel filtration. C: SDS-PAGE gel shows the protein is some degradation product.	176
Figure 80. A: SDS-PAGE gel after nickel from UbS3S4His -BBS5 (1-270). B: The chromatogram of gel filtration purification of UbS3S4His -BBS5. C: The SDS-PAGE gel also confirmed that the product contains very low protein yield.	178
Figure 81. Cartoon diagram of the ubiquitin protein structure (PDB code 1UBI) and the yellow colour represents the loop between strands 3 and 4 where the new peptide was inserted.....	180
Figure 82. A: SDS-PAGE gel of the expression test of UbpepS3S4. B: The chromatogram of the size exclusion chromatography C: SDS-PAGE gel after gel filtration from UbpepS3S4 of the gel filtration.....	181
Figure 83. A: The Nickel column purification chromatogram of USP11 residues (1-244). B: SDS-PAGE gel after Nickel column purification C: Gel filtration chromatogram of USP11 residues (1-244). D: SDS-PAGE gel after gel filtration purification of USP11 residues (1-244).	183
Figure 84. The interaction between USP11 residues (1-244) and Ubpep11S3S4. 200 μ M Ubpep11S3S4 was injected into 20 μ M USP11 residues (1-244) into the sample cell. The ITC graph results showed a binding curve with an estimated Kd value of about 12.2 μ M.	185

List of the tables

Table 1.Oligo deoxy ribonucleotides primers were used in this study.	37
Table 2 Lysis ,Nickel affinity chromatography and gel filtration buffers.	39
Table 3. PCR reaction steps.....	42
Table 4 First PCR reaction for RF cloning procedure.....	45
Table 5. Secondary PCR reaction for RF cloning procedure	45
Table 6 Site direct mutagenesis procedure.....	46
Table 7. Comparison of the yields of 1BKR wild type, 1BKR L17S and 1BKR L17D constructs after gel filtration purification using Superdex 75 HiLoad 16/60 column.	78
Table 8. Results of the crystallisation trials using different crystallisation screens for 30 mg/mL 1BKR-UCHL1 protein after one week	94
Table 9. Results of the crystallisation trials using different crystallisation screens for 30 mg/mL 2GKG-UCHL1 protein after one week	95
Table 10. Results of the crystallisation trials using different crystallisation screens for 30 mg/mL MBP-UCHL1 protein after one week	95
Table 11. Results of crystallisations using different crystal screens for 15 mg/mL 1BKR-UCHL3 protein after one week.....	108
Table 12. Results of crystallisations using different crystal screens for 15 mg/mL 2GKG-UCHL1 protein after one week.....	109
Table 13. Results of the crystallisation trials using different crystal screens for 15 mg/mL MBP-UCHL1 protein after one week.....	109
Table 14.Comparison of yields of 1BKR-UCHL1, 2GKG- UCHL1 and MBP-UCHL1.	112
Table 15.Comparison of yields of 1BKR-UCHL3, 2GG- UCHL3 and MBP-UCHL3.....	113
Table 16. Results of crystallisations using different crystal screens for 20 mg/mL 1BKR-Ataxin (278-291) protein after one week.....	126

Table 17. Results of crystallisations using different crystal screens for 25 mg/mL 2GKG-Ataxin (278-291) protein after one week.....	126
Table 18. Results of crystallisations using different crystal screens for 20 mg/mL MBP-Ataxin (278-291) protein after one week.....	127
Table 19. Results of crystallisation using different crystal screens for 15 mg/mL 1BKR-alpha synuclein (58-79) protein after one week.....	137
Table 20. Results of crystallisation using different crystal screens for 20 mg/mL 2GKG-alpha synuclein (58-79) protein after one week.....	138
Table 21. Results of crystallisation using different crystal screens for 20 mg/mL MBP alpha synuclein protein after one week.....	138
Table 22. Comparison of yields of 1BKR-Ataxin (278-291), 2GKG-Ataxin (278-291) and MBP-Ataxin (278-291) constructs after gel filtration purification.....	140
Table 23. Comparison of yields of MBP- alpha synuclein (58-79) 2GKG- alpha synuclein (58-79) and 1BKR-alpha synuclein (58-79).....	141
Table 24. Results of crystallisation using different crystal screens for 5 mg/mL 1BKR-USP20 ZNFUBP (1-108) protein after one week.....	153
Table 25. Results of crystallisation using different crystal screens for 3.4 mg/mL 2GKG- USP20 ZNFUBP (1-108) protein after one week.....	154
Table 26. Results of crystallisation using different crystal screens for 7.5 mg/mL MBP-USP20ZNFUBP (1-108) protein after one week.....	154
Table 27. Results of crystallisation using different crystal screens for 1 mg/mL 1BKR-BBS5 (1-270) protein after one week.....	164
Table 28. Results of crystallisation using different crystal screens for 1 mg/mL 2GKG -BBS5 (1-270) protein after one week.....	164
Table 29. Comparison of yields of 1BKR-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and MBP-USP20ZNFUBP (1-108) constructs after gel filtration purification.....	167
Table 30. Comparison of yield of 1BKR- BBS5 (1-270) 2GKG- BBS5 (1-270), MBP- BBS5 (1-270) constructs after gel filtration purification.....	168

List of Abbreviations

1BKR	Calponin homology domain of human β -spectrin
2GKG	Receiver domain of frzS
UBHIS3S4	Ubiquitin protein with Histidine tag in the strand 3 and 4
BBS5	Human Bardet Biedl syndrome 5 protein
UCHL1	Ubiquitin carboxyl-terminal hydrolase isozyme L1
UCHL3	Ubiquitin carboxyl-terminal hydrolase isozyme L3
USP20ZNFBP	Ubiquitin specific protease 20 Zinc finger binding protein
Ub	Ubiquitin protein
IMAC	Immobilized metal affinity chromatography
T _m	Melting temperature
RT-PCR machine	Real TIME-Polymerase Chain Reaction machine
PEG	Polyethylene glycol
EDTA	Ethylenediaminetetraacetic acid
GST	Glutathione S transferase protein
TEV	Tobacco Etch virus protease
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
UBpeptideS3S4	Ubiquitin protein with peptide in the strand 3 and 4
MBP	Maltose binding protein
NusA	N-utilization substances tag
TrxA	Thioredoxin tag
DsbA	Disulfide isomerase tag

1. Chapter 1: Introduction

Proteins are essential molecules that play a wide range of different biological functions in all organisms (1). All living organisms synthesise proteins as key components of their biological metabolism and function. Proteins can be classified according to their function such as antibodies, enzymes, structural components, transport and storage proteins (2). Native and recombinant proteins are beneficial to the biopharmaceutical, biotechnological and the agricultural industries (3). Products from these industries benefit the medical industry, human nutrition, the production of detergents, polymers, leather, paper, and plastics (4). Recombinant proteins are currently produced commercially using different system of expression such as bacteria, insect cells, yeast cells and mammalian cells (5). The commercial production of recombinant proteins requires expression and an efficient protein purification process (Figure 1). Generally, recombinant protein production at a large-scale in the milligrams range of pure protein is difficult due to proteins having different physicochemical properties and their natural tendency to aggregate when over-expressed (6).

Maintaining the proteins native structural state is important for biological activity. The native structure is acquired through the process of protein folding (7) (Figure 2), where proteins fold through the interaction of polar residues (hydrogen bonds), oppositely charged residues (ionic interactions) and non-polar residues (hydrophobic interactions)(8). There is consensus that protein folding results from the presence of relatively hydrophobic residues such as alanine, valine and others which should be shielded from an aqueous environment, this is termed as hydrophobic collapse (9). This hydrophobic collapse tends to cause hydrophobic residues to be buried in the core of the

protein structure. The surface of the protein exposed to the solvent is then occupied by polar and charged residues such as asparagine, glutamine, glutamic acid and lysine (10). Improper folding of many proteins has been correlated with pathological diseases such as Parkinson's and Alzheimer's (11). However, it should be noted that many biologically active proteins when in their native state do not adopt a structural fold. These proteins are commonly known as intrinsically disordered proteins (IDPs). IDPs have gained particular interest due to their physiological roles, such as cell signalling and transcription (12).

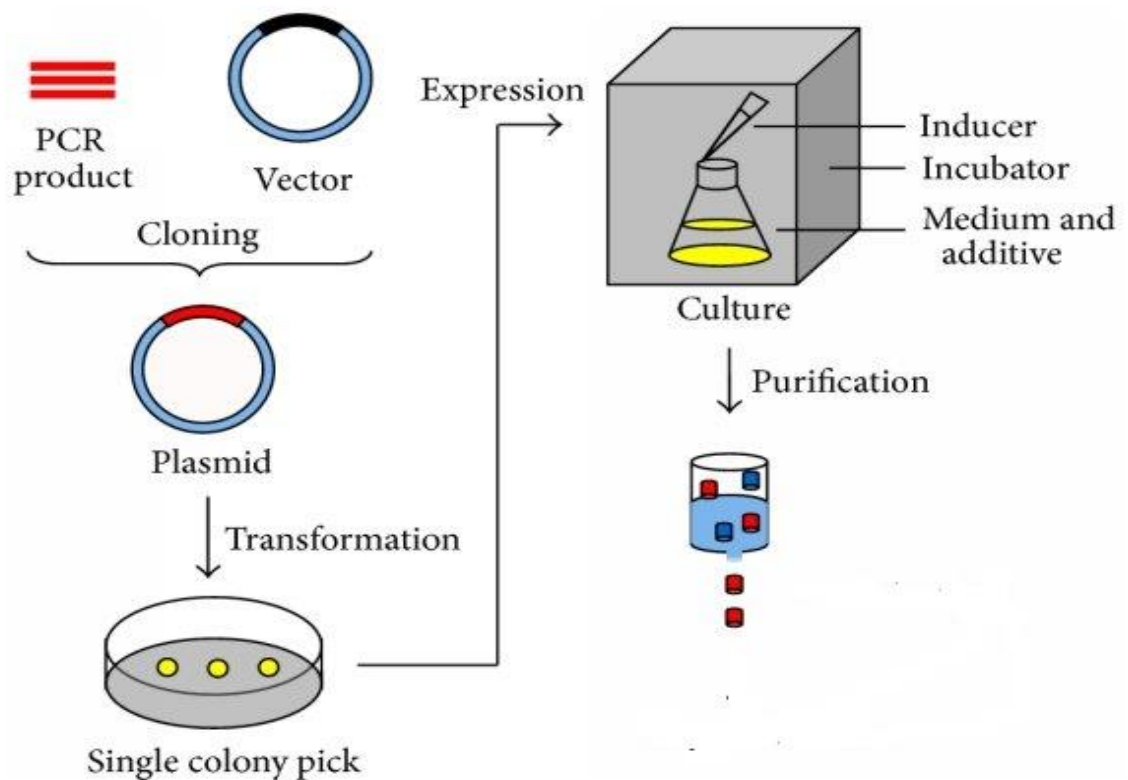


Figure 1. Flowchart showing the process of recombinant protein production started by cloning then expression and purification. Figure adapted from (13).

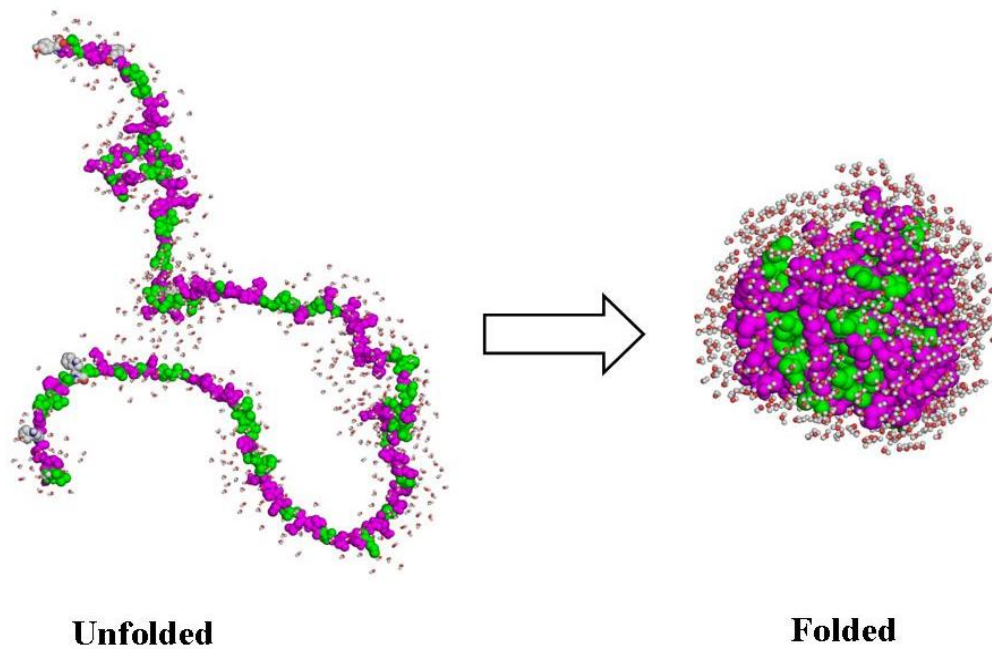


Figure 2. Graphical representation showing the protein folding. During folding, hydrophobic amino acids (represented as green spheres) form the protein core while hydrophilic amino acids (represented as pink spheres) form water exposed surfaces of the protein Figure adapted from (14).

1.1 Challenges and factors affecting the production of recombinant protein

Recombinant protein production from expression systems faces many challenges. For example, the choice of expression host can affect recombinant protein expression, as a bacterial expression system lacks suitable chaperone molecules and post-translational modifications. These render bacterial expression system incompatible for the folding of some eukaryotic origin proteins such as those that harbour disulfide bridges. The lack of suitable pathways in the bacterial host can contribute to the aggregation process (15).

The exposure of purified recombinant proteins towards physical and chemical stresses can affect the stability of the proteins native state by influencing the tertiary structure. This would subsequently lead to an unfolding of the polypeptide chain (16).

Physical conditions such as temperature, pH and protein concentration can induce physical instabilities in the proteins (17) (Figure 3). These physical variables can cause reversible and irreversible protein aggregation. However, a change in pH and temperature may also alter an aggregate from being irreversible to reversible (18).

Proteins are stable and retain their native conformation at a narrow range of pH values, which is a characteristic of the protein. Outside the range of optimal pH, the proteins could potentially undergo transitions from their native conformations which may lead to aggregation (19). Also a change in pH, would redistribute the charge present in the protein and thus affect the attraction or repulsion interactions between molecules. During bio-processing or purification, a recombinant protein may encounter changes in pH which could adversely affect its stability (20).

Temperature plays a crucial role for recombinant protein production, as every protein retains stability below a characteristic specific melting temperature (T_m). When proteins are subjected to conditions above their specific melting temperature, they undergo structural unfolding making them susceptible to aggregation (21). Increasing temperature results in increased frequency of intermolecular collisions and breakage of hydrogen bonds, leading to aggregation (22).

The protein concentration is a significant factor for protein stability as higher protein concentrations increase intermolecular protein-protein interactions, which potentially initiates the process of aggregation (23).

Additionally, chemical changes and the protein structure degradation can be induced by acylation, oxidation, deamination and hydrolysis of peptide bonds (24). Alterations in the chemical structure of a protein lead to a change in polarity and charges, thus influence the attraction or repulsion forces between residues of protein monomers. For example, the presence of unpaired cysteine residues makes them susceptible to form intermolecular disulphide bonds between polypeptide chains and aggregates (25). Tyrosine residues are also known to undergo oxidation to form covalent dityrosine bonds, these chemical changes are generally irreversible in nature (26)

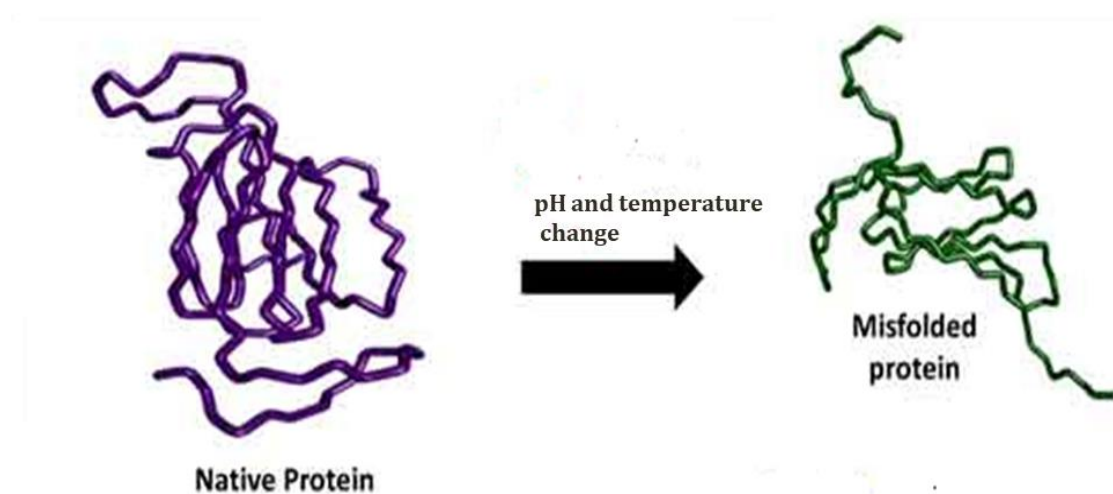


Figure 3.Diagram represents misfolding of native protein under certain circumstances such as pH and temperature change. Figure adapted from (27)

1.2 Characterization of recombinant protein production

Characterization methods for recombinant proteins are wide and diverse. They range from simple visual inspections to sophisticated analytical evaluations that require specialised expertise. The nature and size of protein influences the choice of analytical technique to be utilized. Furthermore, each method is able to assess a certain particular aspect of the protein. These include protein size, morphology or protein concentration (28). Different analytical tools used to detect and characterise protein can be broadly categorized according to chromatographic, light scattering, and spectroscopic methods. These methods differ from each other in terms of their respective detection principles, the detection size ranges and duration of analysis (29).

1.3 Strategies to control stability of the recombinant protein production

1.3.1 Bioinformatical assessment and control of protein aggregation

Computational methods such as *in silico* molecular simulations are currently used to estimate the stability of a specific protein sequence (30). These computational methods are able to identify sequences and motifs which are prone to aggregation. They are valuable in understanding the mechanisms, precursors and inducers underlying the aggregation (31). Recent efforts are aimed at early use of computational tools to identify promising protein molecules which possess desirable conformational stability or those that have a tendency to aggregate through *silico* techniques (32).

1.3.2 Modification of genes of interest by site directed mutagenesis to control aggregation

Site-directed mutagenesis can be used to control protein aggregation. This approach focuses on changing the residues from being hydrophobic to hydrophilic to enhance the solubility of proteins. However, one of the challenges in utilizing a site-directed mutagenesis technique arises when there is limited structural information. This can make it difficult anticipating hydrophobic residues which are useful potential targets for mutagenesis (33).

1.3.3 Expression systems

Both prokaryotic and eukaryotic expression systems are utilised for recombinant protein production. Eukaryotic expression systems are mostly targeted at therapeutic protein production, as these proteins often contain post-translational modifications (34). However, prokaryotic expression systems remain the most desirable choice as they are the cheapest and quickest expression systems.

Bacterial expression systems, including *Escherichia coli* strains are limited by their inability to carry out post-translational modifications (35). This can hinder correct folding of the polypeptide chain. In addition, inclusion bodies inside the cytoplasm of the bacteria will potentially develop as a result (36). Recombinant protein recovery from inclusion bodies in their native conformational state requires solubilisation under denaturing conditions and then to undergo a process of refolding (37). Furthermore, a large number of mutant *E. coli* strains have been designed for recombinant protein expression. One of the commercially available strain is the BL21 *E. coli* strain, which

has been shown to be ideal for routine protein expression. Deficiencies of lon and OmpT proteases in BL21 *E. coli* strain as a result of genetic modifications leads to reduced proteolysis of expressed recombinant proteins and subsequently increased protein stability (38).

1.3.4 Alterations in expression conditions

Several approaches such as a reduction in growth temperature and a change in growth media composition have shown promise in achieving initial soluble expression of recombinant proteins. A reduction in growth temperature is an important factor as higher temperatures increase hydrophobic interactions and thus can increase aggregation during protein expression (39).

The supply of nutrients in growth media is essential for production of properly folded and active recombinant protein. This relationship is dependent on the type of the expressed proteins and requirements for specific co-factors in the composition of the growth media. These specific co-factors may include metal ions such as iron-sulphur as well as polypeptide cofactors (34).

1.3.5 Using fusion tags

One solution used to control protein aggregation and subsequently obtain soluble pure proteins is using fusion tags. Tags which enhance recombinant protein yield, can be classified into two classes. The first class are affinity tags which are used for fast and effective proteins purification. The second class are solubility tags which promote target protein folding and solubility (40). Both classes of tags may be used in combination or

many tags combine these two characteristics in order to overcome aggregation problems and provide an easy route to purification (Figure 4).

Protein expression, respective stability, resistance to proteolytic degradation as well as solubility can be potentially improved through the implementation of fusion tags. Currently, there is a diverse range of fusion tags available ranging from large proteins to small peptides, each with their own inherent specific properties. Despite their use, research is still focused on the production of new fusion tags as the folding and solubility of all target proteins still remains a challenge (41).

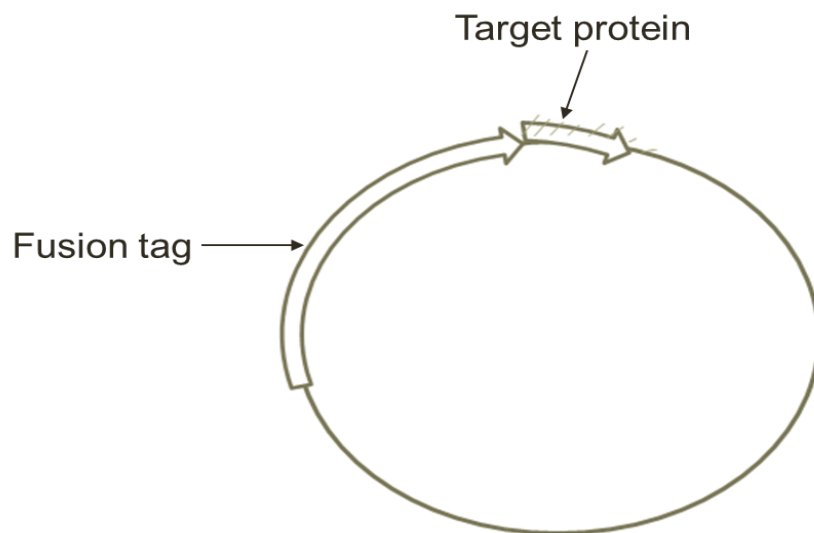


Figure 4 Schematic of a plasmid encoding a target protein and fusion tag.

An ideal solubility tag should improve the expression and purity of the target protein. This is obtained through facilitating folding and enhancing the solubility of fused proteins without any detrimental effect on their respective tertiary structure and biological activity (42). The solubility tag should also be suitable for a number of different proteins and are therefore typically small stable proteins.

1.3.5.1 Mechanisms of solubility tag effects on fusion proteins

Different propositions have been made to explain the mechanism behind how solubility tags improve the solubility of fusion partners.

One mechanism suggests that as a result of the fusion of proteins micelle-like structures are formed (Figure 5). This subsequently results in the separation of target proteins, which are susceptible to aggregation away from the solvent. Simultaneously, the domains of the soluble proteins face outward (43).

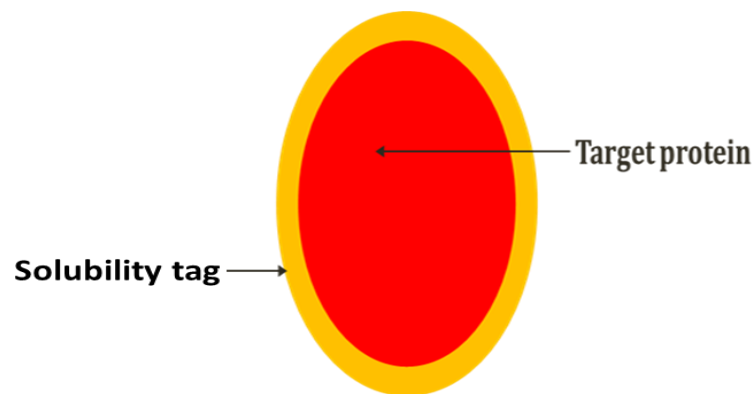


Figure 5. Possible mechanism of solubility tag as micelle. The solubility tag (indicated in yellow) and target protein (indicated in red). During this mechanism, separation of the target protein from the solvent occurs while the domains of the soluble protein simultaneously face outward by micelle-like structures.

A second possible mechanism is the attraction to chaperones by solubility tags. During this mechanism, a protein partner is driven by the fusion tag into a folding pathway that is mediated by a chaperone (44).

A third proposed mechanism is the solubility tags exert intrinsic chaperone-like activity increasing protein folding of the fusion proteins (45) (Figure 6).

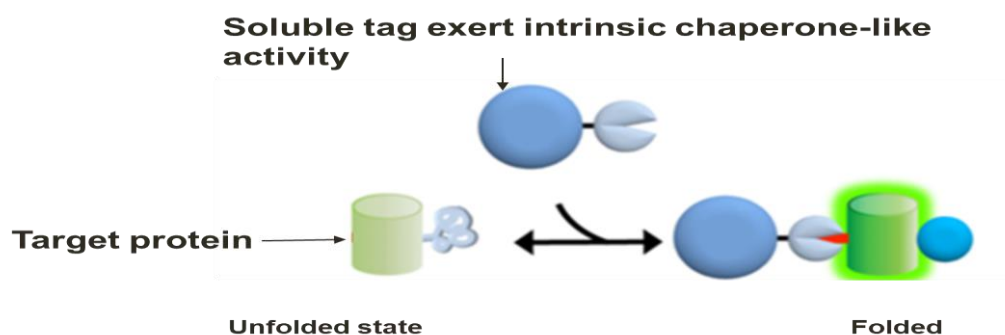


Figure 6. Graphical representation showing that a solubility tag displays intrinsic chaperone activity to protect its fusion protein from aggregation. Figure adopted from (46)

A fourth suggested mechanism is the ability of solubility tags to affect net charges. For example, the highly acidic SET tag inhibits the occurrence of aggregation through repulsive electrostatic interactions and reduces its tendency for globular conformation. It was reported that globular proteins (solubility tags) with an intrinsic high net acidity better enhance the solubility of fusion proteins (47). Different examples of fusion tags and their respective characteristics are discussed below.

1.3.5.2 Examples of existing fusion tags

1.3.5.2.1.1 Poly-histidine tag

A poly-histidine tag is a small tag which consists of typically 6-8 histidine residues and has been commonly used for recombinant protein production through the utilisation of immobilised metal affinity chromatography. The strongest interaction is exhibited between the amino acid histidine and immobilised metal ions such as Cu^{2+} , Zn^{2+} , Co^{2+} and Ni^{2+} . This is a result of electron donor groups present on the histidine imidazole ring which interact with immobilised transition metal ions to form coordination bonds (48). The insertion of the His tag can be performed at the N or C-

terminus, this property allows the use of these tags in all commonly used expression systems e.g. bacteria, mammalian, insects and yeast (49).

1.3.5.2.1.2 Glutathione S-transferase tag

Glutathione S-transferase fusion proteins are the most commonly utilised tool in the field of protein purification (50). The GST tag was first described in 1988 and originates from *Schistosoma japonicum* (51). The purification of fusion proteins can be obtained by using GST tag one-step affinity chromatography methodology based on incubating a crude lysate of the fusion protein with immobilised glutathione. The elution of bound fusion protein can be performed through the addition of 10 mM reduced glutathione in non-denaturing conditions. By protecting fusion proteins from intracellular proteolytic cleavage, the GST tag assists their stabilisation (52). In some cases, due to its large size the GST tag can affect the fusion protein's activity requiring the GST tag to be cleaved from the fusion protein. The cleavage of the GST tag can be carried out by using site-specific proteases such as thrombin or Tobacco Etch virus (TEV) (53). Similar to the His tag, the insertion of GST tag can be at either the N or C-terminus of the protein and thus allowing its utilization in different expression systems such as bacteria, yeast, mammalian cells, and insect cells (54).

1.3.5.2.1.3 Maltose-binding protein (MBP) tag

In 1988, the maltose-binding protein (MPB) tag was first described as a facilitator for peptide expression and purification in *E. coli* (55). Proteins fused with MBP tags can be purified in a one-step affinity chromatography method utilizing amylose resin. The elution of bound fusion proteins uses a buffer containing 10 mM maltose. The MBP tag improves the expression of the fusion proteins by protecting them from being

degraded by the translocation to unique cellular locations containing low protease levels. (56). The insertion of a MBP tag can improve the solubility of fusion proteins in different expression systems including eukaryotic and prokaryotic systems (57). The MBP tag can be inserted at either the N or C-terminus of a protein and is commonly utilised in conjunction with small affinity tags e.g. Histidine tag (58).

1.3.5.2.1.4 NusA, TrxA and DsbA tags

Some other examples of tags which are able to improve the solubility of fusion proteins are N-utilization substances (NusA), Thioredoxin (TrxA) and Disulfide isomerase (DsbA). These tags should be combined with affinity tags such as the Histidine and MBP tags to purify their fusion partners (50). The NusA tag increasing the solubility of fusion proteins as it able to promote hairpin folding and termination (59). In addition, expressed insoluble proteins in *E. coli* becomes soluble when the NusA tag is inserted in the N-terminus (60). Thiorexidin and DsbA tags improve the expression and solubility of fusion proteins by improving their translation efficiency and tertiary structure folding, both tags can be inserted at the N or C-terminus of the fusion proteins (61).

1.3.5.3 Fusion tag applications

A major application of fusion tags is in drug discovery research such as the study of recombinant protein structures (e.g. protein crystallography) and the study of protein-ligand interactions e.g. isothermal titration calorimetry. These techniques requires relatively large amounts of protein in the milligrams range of pure protein (62). Fusion tags provide a high yield of pure proteins while sometimes also assisting in protein crystallography through carrier-assisted crystallisation of fusion proteins. For instance,

the Maltose binding protein (MBP) tag has been used to crystallise many proteins and peptides by promoting the rigidity and compactness of MBP-fused proteins and thus forming crystals (63).

1.4 Background to this study

1.4.1 Novel potential solubility of new suggested tags (2GKG and 1BKR tags)

For the purpose of enhancing the solubility, expression and purification of recombinant proteins new solubility tags were chosen and evaluated as the existing tags cannot be used for all proteins due to different physiochemical properties of the proteins. This requirement is driving a need for new tags, which can be utilised to express various proteins, as the expression and purification of each recombinant protein can be investigated with different tags. For example, the use of large tags e.g. the maltose-binding protein (MBP) can improve the recombinant proteins solubility more than the small tags as it increases the efficient translation of the fusion proteins (64). However, in some cases large tags can interfere with the fusion protein stability and activity during the purification leading to better use of small tags than large tags (65).

When regarding the use of tags for crystallization, some tags can promote crystallization of the fusion proteins more than others such as the presence of the MBP tag has been shown to crystalize many proteins due to its large size and favourable surface properties which improves the crystal packing and promotes the formation of fusion protein crystals (66).

The aim of this study was to identify and evaluate new fusion tags with the aim of improving expression, solubility and crystallization of different proteins. The choice and evaluation of new fusion partners is dependent on different aspects and factors such as the amino acid composition and the size of the solubility tag. The nature of the fusion partner is also important and must be considered, as some fusion partner require large

or small tags. In some cases, the large fusion partner needs small tags better than large tags. The large size tags have wide range of amino acid content exerting biochemical load on the host cell which affect fusion partner stability (67).

The location of tags is also a factor which should be taken into consideration. Depending on their insertion either at the N or C-terminus, different effects on the fusion partners can be gained. Generally, tags placed at the N-terminus are considered to be more advantageous compared to those placed at the C-terminus. This is due to the accurate significance at the initiation sites of translation on the tag from which fusion proteins can gain benefit and hence allow increased efficiency (58).

This study was carried out to identify and choose new solubility tags from the Protein Data Bank (PDB) by employing specific criteria in order to potentially increase the expression yield, solubility, purification and assist in crystallisation of different proteins. We hypothesised that the Protein Data Bank should provide a rich resource for identifying smaller stable protein folds that may be suitable as solubility tags and potentially for carrier driven crystallisation to minimize any impact on a protein's function by steric hindrance. This aim was suggested because Simon Caulton a PhD student in Ingrid Dreveny group made attempts to express, purify and crystallise USP20 domains through the implementation of different tags such as MBP and Thioredoxin. However, the soluble protein yields were low with no crystals being produced with these tags. This created motivation for identification and choice of new solubility tags to improve the expression yield, solubility, purification and assist in crystallisation of the USP20 domains and can be used in the future with other proteins. The PDB was used

to identify new tags as its previously utilized to identify novel solubility tag such as Mocr tag (68).

The criteria that were used to identify new protein tags included:

(1) Proteins of less than 400 amino acids, as the presence of a large tag can interfere with protein function and stability (69).

(2) No disulphide bonds in the protein structures to ensure that the proteins produced under reducing conditions, as reducing conditions are important for proteins production in *E. coli*. This is because *E. coli* doesn't have the ability to assist the efficient stabilization of SH groups or reorganization of correct disulfide bonds which decrease the expression of the target protein (70).

(3) Non-existent ligands in the structure as the presence of ligand binding may change the protein conformation. This would suggest an increased flexibility of the protein which would negatively affect the formation of subsequent crystals (71).

(4) Proteins have high X-ray diffraction resolution data of less than 1.6 Å, as higher resolution structures indicate that the protein is structurally stable with low flexibility and tight packing, which may be beneficial for crystallisation of any fusion proteins (72).

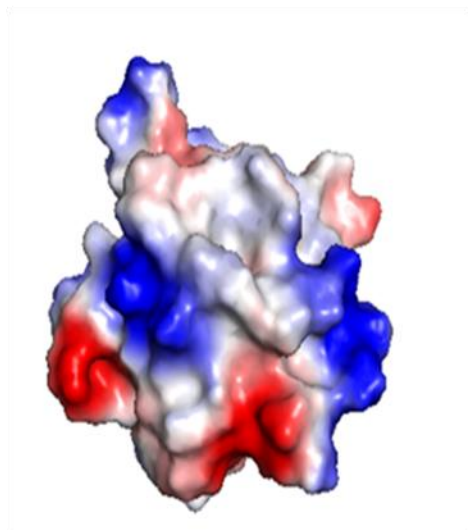
(5) Proteins expressed in the *E. coli* expression system as several fusion proteins use *E. coli* expression system which is the cheapest and fastest system to produce significant amounts of recombinant proteins for crystallization (73).

After applying all the previous criteria, potential fusion partners were 447 initially. Then, the Pymol software and literature reviews were used to further evaluate their characterises by other factors such as globularity, distribution of charges and B-factors values (74) (75).

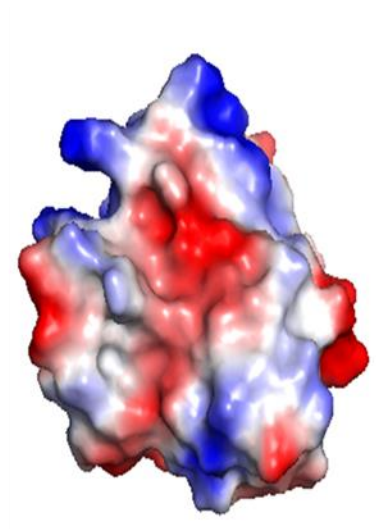
Globularity is important for crystallisation as it decreases conformational flexibility (76). Distribution of surface charges should be relativity even as the presence of large unpolar regions or with a highly positive or negative charge may be more prone to aggregation or interference with protein function (77).

The atoms displacement about its average position is described by a B factor, as higher B values are characteristic of increased movement and higher flexibility which is not favourable for crystallisation (78), as consequence proteins with regions of high B-factors were avoided.

Two of the PDB entries have been selected for use as solubility tags; the human β -spectrin calponin homology domain, (PDB code: 1BKR) (79) and the *Myxococcus xanthus* social motility protein *FrzS* receiver domain, (PDB code: 2GKG) (80). These two tags were selected as they have reasonable surface charge distributions without extremely charged or hydrophobic patches (Figure 7). 1BKR is an alpha-helical protein and the 2GKG consists of a central beta-sheet encapsulated in helices (Figure 8).

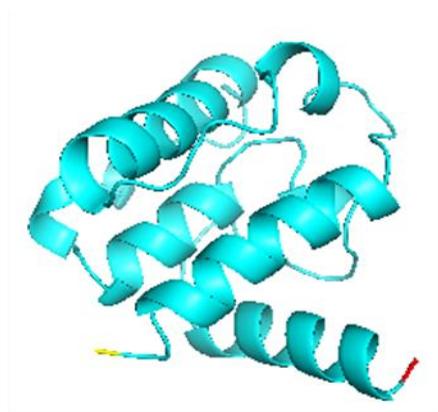


Surface charge distribution of 1BKR

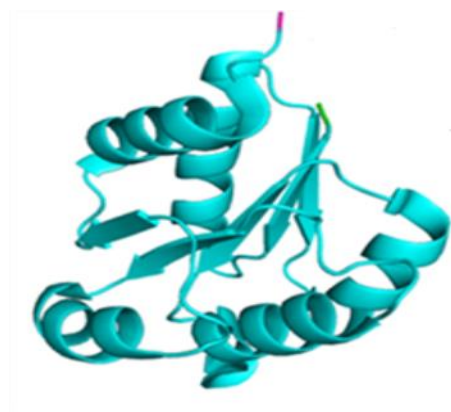


Surface charge distribution of 2GKG

Figure 7. Distribution of 1BKR and 2GKG surface charges. The negative charges are in red colour, the positive charges are in blue colour and the white regions are neutral. (made in Pymol software).



1BKR tag



2GKG tag

Figure 8. Cartoon diagram of the crystal structure of two candidate tags A) 1BKR, B) 2GKG (made in Pymol software)

1.4.2 Chosen model proteins used for investigation with the novel tags

This study investigated these two tags and aimed to evaluate their effectiveness when fused with selected weakly expressed proteins (of unknown structures), highly expressed proteins (known structures) and peptides of known structure which are difficult to express without solubility tags, as the peptides have a high susceptibility for degradation.

Proteins of known structure were chosen in order to study the ability of these novel tags to improve protein expression yield and potentially crystallize these proteins in different space groups (crystal packing) that may be better suited for compound soaking. The chosen proteins were ubiquitin carboxyl-terminal hydrolase isozyme L1 UCHL1 (PDB code 2ETL) (81) and ubiquitin carboxyl-terminal hydrolase isozyme L3 UCHL3 (PDB code 1UCH) (82).

UCHL1 protein is one of the UCHS enzymes, which is a deubiquitinating cysteine protease enzyme that selectively expressed in the brain (83). The UCHL1 protein represents 1-2 percent of the overall brain protein (84). Respective genetic mutations in UCHL1 may be associated with Parkinson disease (85). The crystal structure of UCHL1 enzyme was previously solved at 2.4 Å resolution (Figure 9).

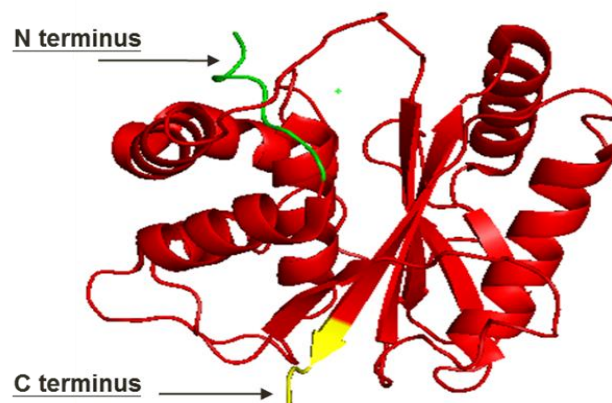


Figure 9. Cartoon diagram of the crystal structure of UCHL1 with the PDB code: 2etl (made using Pymol software). The N terminus represented in green colour while the C terminus represented in yellow colour.

The second model protein was UCHL3 protein, which shares 54% homology with UCHL1 but UCHL3 is different from other UCHs, as it can work as C-terminal hydrolase with both NEDD8 and ubiquitin (86). The crystal structure of UCHL3 protein was solved at 1.8 Å resolution (Figure 10).

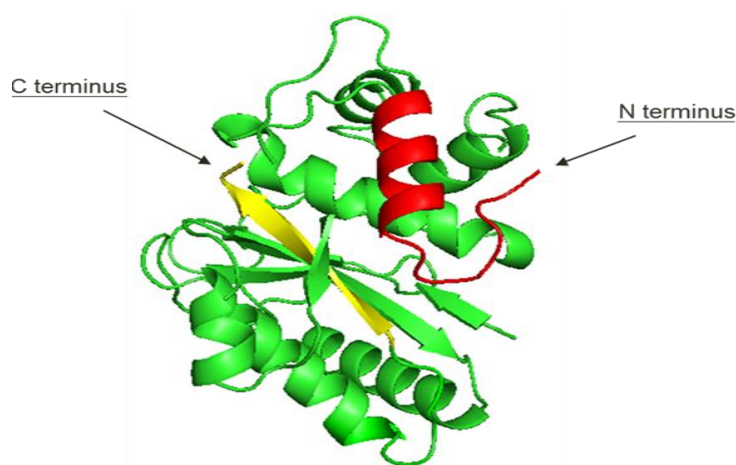


Figure 10. Cartoon diagram of the crystal structure of UCHL3 with the PDB code: 1UCH (made using Pymol software). The N terminus represented in red colour while the C terminus represented in yellow colour.

The second group of model proteins that was investigated were peptides. Producing peptides recombinantly comprising of less than 100 amino acids in *E. coli* is a difficult task due to their susceptibility for degradation (87). The peptides have been selected in order to evaluate whether the new tags that can overcome this challenge. Two peptides from proteins associated with neurodegenerative disease were chosen. These peptides corresponded to ataxin-3 residues (278-291) and α -synuclein residues (58-79). They have been produced previously and their crystal structures were solved via fusion to maltose binding protein (MBP) tag. These two peptides were chosen by specific criteria; (1) the peptides structures were previously solved through fusion with MBP, (2) at high resolution within the range of 1.5 and 2.5Å and (3) were of human origin.

The Ataxin-3 peptide is involved in the pathogenesis of neurodegenerative disease of spinocerebellar ataxia type 3 (88). One such disease defined by the ataxin-3 protein self-aggregation is Machado-Joseph disease (89). The structure of ataxin-3 consists of a globular N-terminal domain described as the Josephin domain. Aggregation of the Josephin domain results in curvilinear protofibrils. The C-terminal unstructured domain, is dynamically disordered and consists of three ubiquitin interacting motifs which are separated by polyglutamine residues (90). As the polyglutamine regions expands to consist of more than 50 residues, a second stage of ataxin-3 aggregation occurs giving the formation of long straight amyloid fibrils (91). The crystal structure of a C-terminal Ataxin-3 peptide (278-291) that comes before the polyQ region was solved at 2.2-Å resolution when fused to MBP tag (92) (Figure 11). The isoelectric point of MBP-ataxin (278-291) is 5.58 with a 42.9 kDa monomeric molecular weight and comprise 392 amino acids. (PDB code: 4YS9).

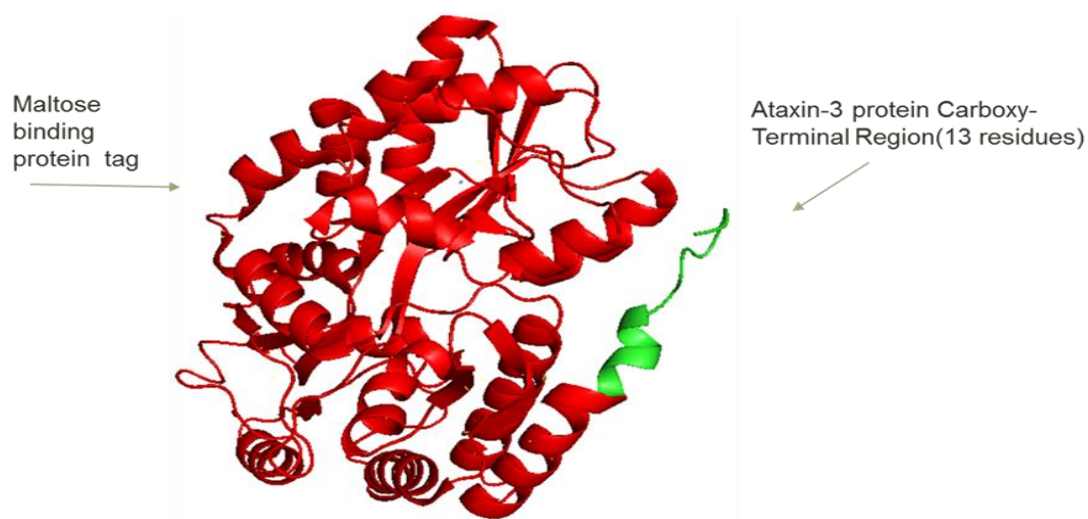


Figure 11. Cartoon diagram of the crystal structure of MBP-ataxin (278-291) PDB code: 4YS9 (made using pymol software). The MBP tag represented in red colour while ataxin (278-291) represented in green colour.

The second chosen model peptide was α -synuclein (residues 58-79) peptide, which is a characteristic feature of Parkinson's disease. The Lewy bodies which present in Parkinson's disease is mainly made up of α -synuclein (residues 58-79) peptide (93). The aggregation of the peptide forms synucleinopathies, which were found to be present in various human neurodegenerative diseases (94). The mechanisms behind the conversion of α -synuclein in its functional native state to the aggregates of amyloid are still unexplained (95). The crystal structure of the α -synuclein peptide (58-79) peptide was solved by fusion with maltose binding protein(MBP) tag at 1.6 Å resolution (96) (Figure 12) (PDB code: 3Q28).

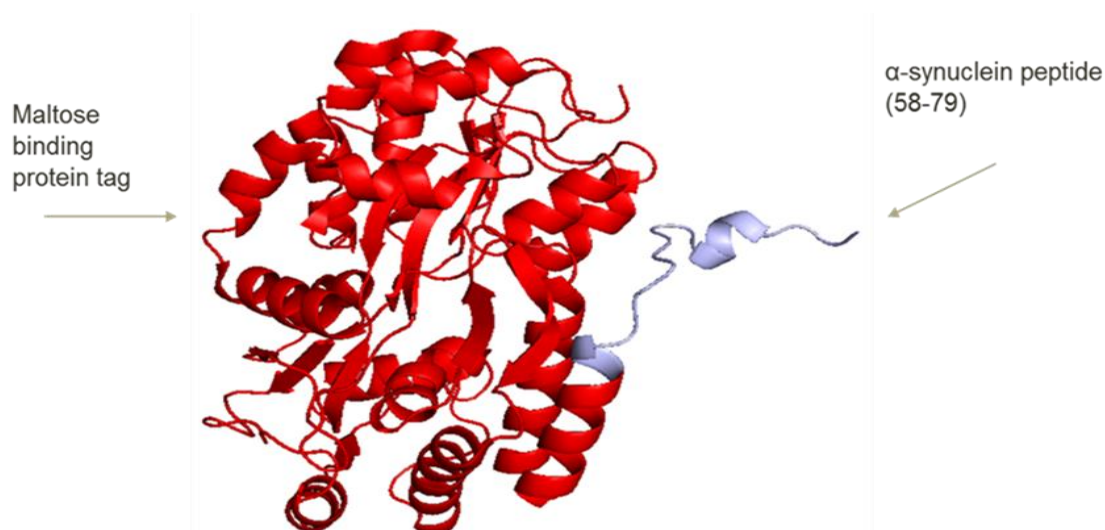


Figure 12. Cartoon diagram of the crystal structure of MBP- α - synuclein (58-79) PDB code: 3Q28 (made using Pymol software). The MBP tag represented in red colour while α -synuclein peptide (58-79) represented in violet colour.

The third group investigated were model proteins that were difficult to express and purify, whose structures have not yet been solved, to study the efficacy of the novel tags in terms of expression, solubility and crystallization when fused with these model proteins. The chosen proteins were a predicted zinc finger domain of ubiquitin specific protease 20 USP20ZNFUBP (1-108) and the predicted pleckstrin homology domains (PH domains) of Bardet-Biedl syndrome 5 protein BBS5 (1-270).

The ubiquitin specific protease 20 (USP20) is a deubiquitinating enzyme, which is implicated in various cellular pathways, such as the regulation of DNA-damage response pathway, as well as checkpoints in the cell cycle (97). USP20ZNFUBP (1-108) is a zinc finger domain which is part of USP20 protein. Zinc finger domains are functional and generally fold independently, the structure is stabilized by coordination with zinc ions via histidine and cysteine residues (98). Zinc finger domains in proteins are structurally diverse and have important roles in DNA to protein interactions and

membrane association (99). The isoelectric point of USP20 ZNFUBP (1-108) is 6.39 with a monomeric molecular weight of 12 kDa.

The second model protein was the Bardet-Biedl syndrome 5 protein BBS5 PH domains (residues 1-270) that is membrane-associated. BBS5 is a protein encoded by the *BBS5* gene which is directly linked to Bardet-Biedl syndrome (100). The disease is a result of genetic mutations in the gene *BBS5* and is characterised by severe pigmentary retinopathy, kidney malformation and mental illness. Also, Bardet-Biedl syndrome patients may have diabetes mellitus, hypertension and congenital heart disease (101) (102). The isoelectric point of BBS5 (1-270) is 8.9 with a monomeric molecular weight of 30.7 kDa.

1.4.3 Design and investigate the suitability of the ubiquitin protein for protein affinity, solubility and protein scaffolds

In addition, to the candidate tags identified from the Protein Data Bank a well characterised highly thermostable protein was also investigated for improvement of the expression and purification of difficult to express proteins. Currently fusion tags that improve the expression and solubility cannot be used for all proteins. Thus, novel small tags that are unlikely to interfere with a protein function are required. This is driving the research into the discovery of new tags. We have investigated the ubiquitin protein for its suitability as a fusion tag. This protein has high stability and solubility which suggest it would be suitable as a fusion tag, as many studies have shown that the protein stability and solubility are important criteria for a fusion tag when looking at improving expression, solubility and crystallization of the fusion proteins (103). Previous research has shown that a Small Ubiquitin-like Modifier (SUMO) tag when fused with different proteins and peptides had a comparable performance with many commonly used affinity tags including glutathione S-transferase and maltose-binding protein (MBP) (104) (56) (105). A SUMO tag exhibited similar performance to that observed for MBP and GST tags with different proteins such as matrix metalloprotease-13 (MMP13) and green fluorescent protein. This result indicated that the SUMO tag can be used as fusion tag due to its high solubility and stability (106). Ubiquitin protein has also been shown to be capable of being used as a fusion tag as it can enhance the fusion partner's translation effectiveness (58), Therefore, we assumed that ubiquitin protein can be used as fusion tag, as it is one of the more stable and soluble protein which can be used as affinity and solubility tag.

A second aim of this research was to determine whether the ubiquitin protein can act as protein scaffolds for the insertion of large peptides, which can then be used for high affinity target binding due to its high stability and solubility. Previous research has shown that this protein can be used for high affinity target binding and it has been shown to be a well suited scaffold protein which can be adapted to specific tasks as therapeutic purposes (107).

In this study, we have engineered the ubiquitin protein so it could be used as an affinity tag, solubility tag and tested the suitability of this protein as a protein scaffold for high affinity target binding via a large extension of a loop region with a peptide.

1.4.3.1 Ubiquitin as affinity, solubility tag

With the aim of producing a new enhancing solubility tag, a loop region of the ubiquitin protein was modified according to specific design (Figure 13). Ubiquitin as a highly stable and soluble protein is expected to promote the solubility and expression of many target fusion proteins (108). To unfold ubiquitin temperatures above 90°C are needed (109). The Ubiquitin protein as a fusion tag was investigated with chosen model proteins, which are weakly expressed, in order to hopefully see an improvement in terms of protein yield.

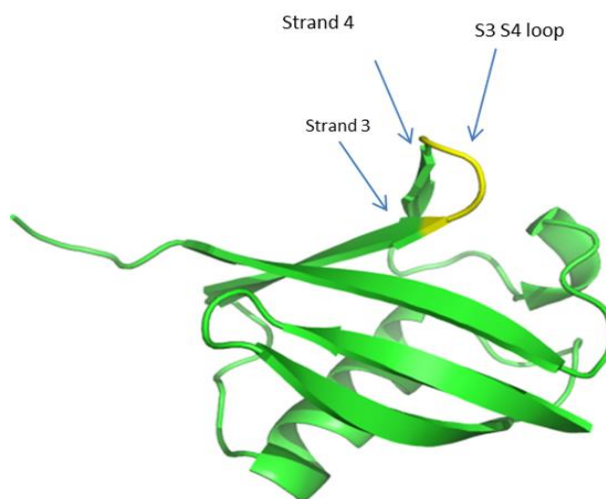


Figure 13. Cartoon diagram of the crystal structure of ubiquitin protein (PDB code 1UBI). The figure represents the loop between strands 3 and strand 4 in yellow colour (made using Pymol software).

1.4.3.2 Ubiquitin as a protein scaffold

A vast range of biotechnological applications such as research, diagnostics and therapy have implemented engineered protein scaffolds as a valuable tool (110). These protein scaffolds have favourable properties, including a small size and the possibility to be conjugated and modified with diverse sequences that can then interact with a variety of protein targets with high affinity (111).

It has been shown that the next-generation phage display technology is useful for efficient selection of target-specific peptides which bind to the deubiquitinating enzyme USP11 residues 1-244 (112). USP11 has been implicated in cancer (DNA damage repair by homologous recombination) and viral-host interactions (e.g. human papilloma virus). Consequently, it is considered a promising drug target. Therefore, peptides that specifically bind USP11 to compete with substrates or other binding partners may form the basis of the development of peptidomimetic drugs (113).

In this study, the ubiquitin protein was used to incorporate a constrained peptide that has been identified as interacting with the USP11 N-terminal DUSP-UBL domains from a S-S constrained phage display peptide library into one of ubiquitin's loop regions between strands 3 and 4. Incorporation of this region into the ubiquitin scaffold was investigated by isothermal titration calorimetry technique with USP11 residues 1-244 (a region that does not interact with ubiquitin).

1.5 The aims of this study

The overall aims of this study were to:

1.1 Investigate two novel tags identified from the PDB, with codes: 1BKR (calponin homology domain from human β -spectrin) and 2GKG (*Myxococcus xanthus* social motility protein *FrzS*), to evaluate if they can be used as solubility tags and if they are suitable for crystallization studies by acting as carrier proteins. These tags were expressed, purified and crystallized using different chosen model proteins and peptides.

1.2 Compare the effectiveness of these two novel tags with the effectiveness of the maltose binding protein (MBP) tag, when fused with chosen proteins with specific criteria as discussed previously. The MBP tag was used in our study as it is commonly used as a reference soluble tag in many studies with novel tags (114).

2. Design and investigate the suitability of the ubiquitin protein as affinity and solubility tag in terms of improving solubility when fused with difficult to express proteins such as the membrane associated BBS5 and USP20ZNFUBP, which have a tendency to aggregate.

3. Investigate the suitability of ubiquitin protein as a protein scaffold allowing incorporation of long peptides into loop regions of the protein between strands 3 and strand 4.

1.6 Specific aims

For design and evaluation of two novel tags 1BKR and 2GKG and if they can be used as solubility tags:

- To reclone the sequences of 1BKR and 2GKG into expression vectors with an N-terminal affinity tag (Histidine tag) followed by the multiple cloning site (MCS) such that, the two tags can be fused to proteins of interest with a linker of three alanine residues to limit the linker length which should aid crystallization of the fusion proteins (115).
- To evaluate the expression, purification, stability and solubility of the 1BKR and 2GKG proteins **without any fusion** of target proteins to confirm that these proteins have high expression levels, high solubility and stability in isolation (116). Then to also test the crystallisability of the tags alone.
- To evaluate the expression, purification and crystallization of the novel tags **fused** with different chosen model proteins and peptides.

For comparison, the effectiveness of the two novel tags was to be compared with the effectiveness of the maltose binding protein (MBP), when fused to selected proteins:

- To initially compare the tags in terms of expression, solubility and purification using small scale 100 mL bacterial cultures under auto-induction conditions. The small scale cultures using qualitative SDS-PAGE as an indicator of the expected yields of proteins produced at a large scale. The reasoning for auto-induction as a basis of comparison is its ability to screen efficiently the expression and solubility of various clones simultaneously (117).

- To then compare the same proteins as above in a large scale 2 litres bacterial culture and using isopropyl β -D-1-thiogalactopyranoside (IPTG) induction. This allows comparisons to be made quantitatively and facilitates high-yields of soluble proteins for subsequent crystallisation trials for each model protein.

For design and investigation of the suitability of ubiquitin protein, as a protein affinity and solubility tag:

- To insert a 6xHistidine tag as affinity tag into the ubiquitin protein loop between strands 3 and strand 4 (UbHIS3S4) that is highly variable in length and sequence amongst proteins adopting an ubiquitin-like fold.
- To evaluate the expression and purification levels of the UbHIS3S4 protein without any fusion protein.
- To fuse UbHIS3S4 protein with two difficult to express and purify proteins (BBS5 and USP20ZNFUBP) and evaluate the expression, purification and crystallization of these two fusion proteins

For the investigation of the suitability of ubiquitin protein as a protein scaffold for the incorporation of long peptides into loop regions of the protein:

- To clone the peptide of interest into the loop region of the ubiquitin protein between strands 3 and strand 4 (UBpeptideS3S4)
- To express and purify the UBpeptideS3S4 protein in 2 litres bacterial culture to obtain a high yield for interaction studies using isothermal titration calorimetry technique
- To express and purify the USP11 N-terminal region (residues 1-244) in 2 litres bacterial culture to obtain a high yield for isothermal titration calorimetry technique

- To use isothermal titration calorimetry technique for interaction studies between USP11 (residues 1-244) and the ubiquitin protein with the identified peptide insertion (UBpeptideS3S4).

2. Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Chemicals

Sigma-Aldrich (Darmstadt, Germany) supplied most of the chemicals used in this research. Molecular Dimensions supplied the crystallization screens.

2.1.2 Enzymes

All the enzymes were provided by New England Biolabs, e.g. *NotI*, *HindIII*, *XhoI*, *BamHI*, alkaline phosphatase, *DpnI*, T4 DNA polymerase, T4 DNA ligase, Pfu polymerase and alkaline phosphatase.

2.1.3 Bacterial Strains

Nova Blue and B21(DE3) *Escherichia coli* strains were used for this study. Nova Blue cells have been used for cloning and site-directed mutagenesis experiments, while for protein expression *E. coli* B21(DE3) cells have been used.

2.1.4 Plasmids

For this study, the expression vector used for this research was pRSF13 vector, this vector includes a lactose operon controlled by a promoter of T7 RNA. The vector pRSF13 includes an open reading frame encoding resistant to antibiotics for kanamycin resistance. The restriction sites were designed to follow insertion of either the 1BKR or 2GKG genes. The histidine tag, solubility tag and target gene were inserted at multiple cloning sites. The multiple cloning sites consists of eight sites used for in-frame cloning,

they are in this order 5'-3' *NotI*, *NheI*, *KpnI*, *BamHI*, *EcoRI*, *SacI*, *HindIII* and *XhoI*. Cloning into the first two *Not I* or *NheI* restriction enzyme sites resulted in a fusion containing a three-alanine linker called the fixed arm linker.

All the oligonucleotides used for PCR amplification, site- directed mutagenesis, free cloning restrictions sites and sequencing are listed in table 1. Highlighted in green in the table are the bases which correspond to the restriction enzymes used for ligation and cloning purposes. All oligonucleotide primers were synthesised by Sigma-Aldrich (Damstardt, Germany).

Oligo primers namw	Forward primers (5' to 3')	Reverse primers (3' to 5')
1BKR mcs L 17 D	GGTTTTCATCTGGCACCAATCC AGCAGCGCATCTTTTCGC	GCGAAAGATGCGCTGCTGGATTGGTG CCAGATGAAAA CC
1BKR mcs L 17 S	GTTTTCATCTGGCACCAACGACA GCA GCGCATCTTTTCGC	GCGAAAGATGCGCTGCTGTCGTGGTG CCAGATGAAAA C
Ubiquitin change the loop between strand 3 and 4	AATTCTCGAGGTGGA GCCCAGT GACACCATCG	TTCCAA GCTTTCACCCA CCTCTGAGA CGGAGTACCAG
Insertion His tag in the Ubiquitin between strand 3 and 4	CGCGGA TCCCA CCAACCACCAC CACCAC	CTC AGA GGTGGGGAATTC GGT ACC GCT AGC AGT ACT GCA TGC TGA
Insertion peptide in the Ubiquitin protein between strand 3 and 4	GA TCCCTGGCCA GCCTGCTGCC CAGGAGGAGC G	GA TCC GCT CCTCCT GGG CAG CAG GCT GGC CAG G
USP20 ZNFUBP (1-108) - 1BKR – pRSF13	ATGCGGCCGCAATGGGGGACT CCAGGGACC	GCGG CTCGAG TTA AGA GAA CTT GGA AGA GGA GCC C
USP20 ZNFUBP (1-108) -2GKG – pRSF13	ATGCGGCCGCAATGGGGGACT CCAGGGACC	GCGG CTCGAG TTA AGA GAA CTT GGA AGA GGA GCC C
USP20 ZNFUBP (1-108) -MBP – pRSF13	GGA GATA TACCA TGCGCATC ATCA TCA TCA TCATATGAAAA TCGAA GAAGGTAAACTGG	AAAGGTCCCTGGAGTCCCCCATTTGCG GCCGCATTAGTCTG
USP20 ZNFUBP (1-108) –UBHISS3S4 – pRSF13	ATTGCTAGCATGGGGGACTCCA GGGAC	ATA AAGCTT TTA AGA GAA CTT GGA AGA GGA GCC C
BBS5(1-270) -1BKR – pRSF13	ATGCGGCCGCAATGTCGGTGCT GGA TCGCTTTG	GCGG CTCGAG TTA TTC CAT CTC ATA ATC AAC TCC
1BKR - human alpha-synuclein (58-79)	GCCGCaAGCAAGACCAAGGAG CAGGTGA CCAACGTGGGCGGC GCCGTGGTGACCTAA G	GATCCTTAGGTCACCACGGCGCCGCC CACGTTGGTCACCTGCTCCTTGGTCTT GCTrGC
2GKG - human alpha-synuclein (58-79)	GGCCGCaAGCAA GACCAAGGA GCA GGTGA CCAACGTGGGCGG CGCCGTGGTGACCTAA G	GATCCTTAGGTCACCACGGCGCCGCC CACGTTGGTCACCTGCTCCTTGGTCTT GCTrGC
MBP - human alpha-synuclein (58-79)	GGA GATA TACCA TGCGCATC ATCA TCA TCA TCATATGAAAA TCGAA GAAGGTAAACTGG	TCA CCTGCTCCTTGGTCTTGCTTGC GG CCGCATTAGTCTG
1BKR -Ataxin-3 Carboxy- Terminal Region	GGCCGCaAGCGAGGA GCTGA GG AAGAGGAGGGAGGCCTA CTTC GAGTAA G	GATCCTTACTCGAA GTAGGCCTCCCT CCTCTTCCTCAGCTCCTCGCTTGC
2GKG-A taxin-3 Carboxy- Terminal Region	GGCCGCaAGCGAGGA GCTGA GG AAGAGGAGGGAGGCCTA CTTC GAGTAA G	GATCCTTACTCGAA GTAGGCCTCCCT CCTCTTCCTCAGCTCCTCGCTTGC
MBP-Ataxin-3 Carboxy- Terminal Region	GGA GATA TACCA TGCGCATC ATCA TCA TCA TCATATGAAAA TCGAA GAAGGTAAACTGG	TCCTCTTCCTCAGCTCCTCGCTTGC GG CCGCATTAGTCTG
1BKR-UCHL1	ATGCGGCCGCAATGCAGCTCA AGCCGATGGAGATC	GCGG CTCGAG TTA GGC TGC CTT GCA GAG AGC
2GKG-UCHL1	ATGCGGCCGCAATGCAGCTCA AGCCGATGGAGATC	GCGG CTCGAG TTA GGC TGC CTT GCAGAGAGC
MBP-UCHL1	GGA GATA TACCA TGCGCATC ATCA TCA TCA TCATATGAAAA TCGAA GAAGGTAAACTGG	TCTCCATCGGCTTGAGCTGCATTGCG GCCGCATTAGTCTG
1BKR-UCHL3	ATGCGGCCGCAATGGAGGGTC AACGCTGGCTG	AGC AAGCTT CTA TGC TGC AGA AAG AGCAAT CGC
2GKG-UCHL3	ATGCGGCCGCAATGGAGGGTC AACGCTGGCTG	AGC AAGCTT CTA TGC TGC AGA AAG AGCAAT CGC
MBP-UCHL3	GGA GATA TACCA TGCGCATC ATCA TCA TCA TCATATGAAAA TCGAA GAAGGTAAACTGG	GCAGCCAGCGTTGACCTCCATTGCG GCCGCATTAGTCTG

Table 1.Oligo deoxy ribonucleotides primers were used in this study.

2.1.5 DNA ladders

100 base pairs and 1 Kilo base DNA size ladders utilized in this study were supplied by NEB New England Bio lab.

2.1.6 Media

Luria-Bertani (LB) medium 10 g/L tryptone, 5 g/L yeast extract and 10 g/L sodium chloride was provided as a dried powder by Sigma-Aldrich.

2.1.7 Media for auto induction

Per 500 mL media: 5 g of N-Z amine, 2.5 g yeast extract and 1 mL MgSO_4 of 1 M solution were used.

Salts: For each 500 mL of culture, these solutions were added 12.5 mL of 1 M solution Na_2HPO_4 , 12.5 mL of 1 M solution KH_2PO_4 , 25 mL of 1 M solution NH_4Cl , and 2.5 mL of 1 M solution Na_2SO_4 .

Sugars consist of Glycerol (100%) 62.5 mL, D-Glucose 1.12 g, a-Lactose (monohydrate) 9 g and top up with autoclaved distilled water to 250 mL.

2.1.8 Cell strains

For storage of cell strains, a glycerol stock solution consisting of 50% (v/v) glycerol and LB medium was used.

2.1.9 Preparation of agarose gel

Agarose gel was made by 0.4 grams of agarose in 40 mL of 1x TAE (40mM Tris, 20mM Acetate and 1mM EDTA and typically has a pH around 8.6) solution. After the 1% (w/v) of agarose gels was made, the melting agarose was transferred in a gel plastic box and mixed with 5 μ L of NANCY-520 stain and a suitable comb at the top end was inserted.

2.1.10 Minipreparation Buffers plasmid for DNA isolation

- Resuspension solution consist of 10 mM EDTA ,50 mM Tris-HCl, pH 8.0 and 100 mg/mL RNase A.
- lysis solution: 200 mM sodium hydroxide and 1% (w/v) sodium dodecyl sulfate
- Neutralization buffer: 3 M potassium acetate, pH 5.5.
- Wash buffer: 1 M sodium chloride,50 mM of 3-(N-morpholino) propanesulfonic acid) buffer, pH 7.0, 15% (v/v) 2-propanol.
- Elution solution: autoclaved distilled water

2.1.11 Buffers for protein purification

Buffers have been used in all purification steps shown in table 2.

Lysis Buffer	Immobilized metal affinity chromatography buffer	Gel filtration Buffer
• 20 mM Tris • 200 mM sodium chloride • 50 mM Imidazole • 3% glycerol • pH 7.5	• 20 mM Tris • 200 mM sodium chloride • 500 mM Imidazole • 3% glycerol • pH 7.5	• 20 mM Tris • 200mM sodium chloride • 1 % glycerol • pH 7.5

Table 2 Lysis ,Nickel affinity chromatography and gel filtration buffers.

2.1.12 Crystallisation solutions

Crystallisation screens were bought from Hampton Research (Morphius, PEG, Screen). Handmade grid screens were mixed and ranged from low pH till high pH, different concentrations of PEG and ammonium sulfate. The PEG screen had a starting PEG concentration of between 400 to 6000 and the pH ranged from 4.5 to 9.5. The ammonium sulfate concentrations ranged between 0.5 M to 2.5 M and pH ranged from 4.5 to 9.5.

2.2 Methods

2.2.1 Bacteria strains growth

In autoclaved LB medium, *E. coli* was grown at 37°C with a suitable antibiotic. The antibiotic concentrations used were: Chloramphenicol 35 µg/mL, Ampicillin 100 µg/mL and Kanamycin 50 µg/mL (118).

2.2.2 Transformation of competent *E coli* strains

The transformation was performed by adding 50 µg DNA into competent cells. The suspension of cell / DNA was incubated on ice for 10 minutes, then the suspension was heat-shocked at 42°C for 1 minute and re-incubated on ice for 2 minutes. Then 500 µL of LB was added to the cell/DNA mixture and incubated for 1 hour in a shaking incubator at 37°C. Up to 200 µL of this mixture was spread on LB agar plate containing the suitable antibiotic and incubated overnight at 37°C (119).

2.2.3 Measurement of DNA concentration

A NanoDrop™ 2000/2000c Spectrophotometers machine (Thermo Scientific, MA, US) have been used to measure the nucleic acids concentration at 260 nm absorbance (120).

2.2.4 DNA plasmid isolation (mini-preparation):

A Plasmid Miniprep kit (QIAquick, company) was used to perform the DNA plasmid isolation protocol. The first step was 10 minutes centrifugation of 5 mL overnight culture at 5000 RPM. The pellet was re-suspended in 100 µL of re-suspension buffer. Then, 250 µL lysis buffer was added to lyse the cells. The lysis was stopped with

150 µL of neutralization buffer. To remove the cell debris and precipitated protein after neutralization buffer, the lysate was centrifuged for 10 minutes at 5000 RPM the supernatant transferred into a QIAquick column. These columns have matrix which can bind to different sizes of DNA molecules. The centrifuging was done to bind DNA molecules to the columns at 12000 RPM for 2 minutes. Then 750 µL of wash solution was added to remove the impurities out of the column. Finally, 50 µL of autoclaved distilled water was used for elution (121).

2.2.5 Molecular cloning methods

There were three different cloning methods used in this study as discussed below.

2.2.5.1 Cloning by using restriction sites

Cloning by the use of restriction enzymes and restriction sites was done by the following steps: Polymerase Chain Reaction, DNA Purification from agarose gel, digestion of DNA by restriction endonucleases and DNA ligation. The following sections explain all these steps in detail.

2.2.5.1.1 Polymerase Chain Reaction (PCR)

The first step of the cloning was using template DNA molecule to amplify specific DNA fragment using a PCR thermocycler T100™ Thermal Cycler (Bio-Rad Herts, UK).

Each PCR reaction consisted of a master-mix, forward primer, reverse primer, 50% DMSO, distilled water and 0.5 µL PHUSION polymerase. The master mix was composed of DNA template (150 ng), 22.5 µL 5X HF buffer, 1.8 µL dNTPs, 56.25 µL

distilled water, respectively. Reactions were performed as follows started by 95°C, 1 min and an annealing temperature 55-67°C, 1 min with an extension temperature was 72°C, 1 min per 1000 base pair, and repeated 20-25x. (122). Table 3 shows the PCR protocol used.

	Step	Temperature	Time
	First denaturation	95 °C	5 minutes
24 x	Denaturation	95 °C	30 seconds
	Annealing	55-67 °C	1 min
	Extension	72 °C	1 min per 1 kbp
	Final extension	72 °C	10 min
	End of reaction	4 °C	∞

Table 3. PCR reaction steps

2.2.5.1.2 DNA Purification from agarose gel

At 100 Volt for 30 minutes, the gel was electrophoresed. Under Dark Reader Transilluminator UV light the DNA bands were visualized using NANCY-520 stain (123).

The DNA plasmid or PCR products were purified from agarose gels using PCR cleanup and Gel Extraction Qiagen kits. A clean scalpel was used to cut out a DNA band of interest from a 1% (w/v) agarose gels and placed into a sterile Eppendorf tube. Then solubilizing solution was added for 10 min at 60°C to dissolve the agarose. After agarose was dissolved, the mixture of DNA and melted agarose transferred into a QIA-quick

column. To make the DNA molecule bind to the column, samples were centrifuged at 13000 RPM for 2 minutes. Then, 750 µL of wash solution was added to the column. DNA elution was carried out with 50 µL of distilled autoclaved water (124)

2.2.5.1.3 Restriction enzyme digest of DNA by restriction endonucleases

For PCR products digestion reactions, 1 µL of the prescribed buffer was added to 7 µL of PCR product and 0.5 µL of each restriction enzyme, then incubated at room temperature overnight. The digestion reaction of vector for gene insertion was done in the same manner as the PCR product digestion but it had additionally step which was de-phosphorylation. The de-phosphorylation was done by adding 1 µL of Antarctic Phosphatase and 2 µL of Antarctic Phosphatase buffer to the digestion reaction and incubated for 2 hours at 37 ° C (125).

2.2.5.1.4 DNA ligation

T4 DNA Ligase enzyme was used to perform the ligation reaction. The ratio of the ligase reactions was performed at 6:1 ratio of insert volume to vector volume. The ratio volumes of DNA required was calculated by the following equation:

$$\text{Volume of insert} = \frac{\text{concentration of vector} \times \text{kb size of insert}}{\frac{\text{kb of vector}}{\text{Concentration of insert}}}$$

T4 DNA Ligase Buffer was then added to the DNA insert/ vector mix and the reactions were incubated at 16°C overnight (126).

2.2.5.2 Cloning using Annealing Oligonucleotides Primers

Second method of Molecular cloning was used Annealing Oligonucleotides primers. This approach is ideal for peptide cloning as it makes an artificial DNA fragment of up to 80 base pairs and can be cloned into a vector of choice. The primers of the artificial DNA fragments were made by overhanging nucleotides in both forward and reverse primers (104) (Figure 14). Both primers were mixed and formed double strand DNA and directly ligated into the multiple cloning sites of the digested vector. Both forward and reverse primers were annealed at 90°C for 10 minutes.

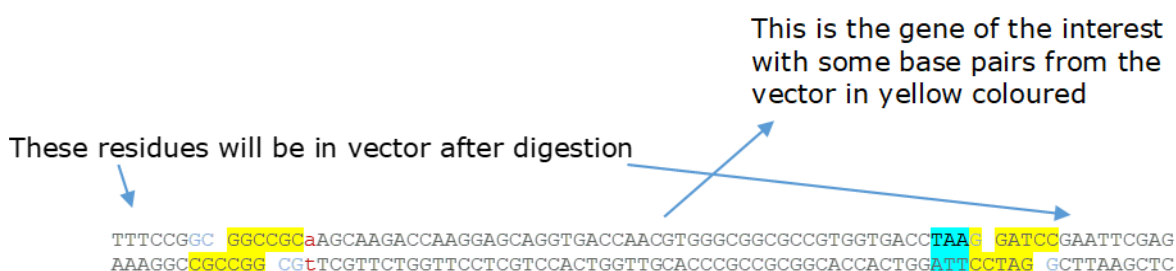


Figure 14. Example of the primer design of Annealing Oligonucleotides Primers.

2.2.5.3 Cloning using Restriction-free (RF) method.

In this protocol, the DNA fragment can be inserted in the vector of choice without restriction sites by specific design of the primers. The primers had complementary sequences for both vector of choice and desired insert, these primers are used in the first PCR reaction. The first PCR reaction was prepared as normal as PCR reaction as discussed before. The mega primers were the results product after the first PCR reaction and the product is purified. The vector of choice then acts as template along with mega-primers in the secondary PCR reaction. In both directions, the vector is amplified and a nicked hybrid molecule formed by mega-primers which allows the complementary

strands of the vector to anneal. To degrade the parental vector, the *DpnI* enzyme is used and the final product is transformed into competent cells (127). The first and second PCR reactions steps are shown in tables 4 and 5.

	Step	Temperature	Time
	First denaturation	98°C	30sec
35 x	Denaturation	98°C	8sec
	Annealing	55-60°C	20sec
	Extension	72°C	15-30sec/kb
	Final extension	72°C	5min
	End of reaction	4 °C	∞

Table 4 First PCR reaction for RF cloning procedure

	Step	Temperature	Time
	First denaturation	98°C	30sec
15 x	Denaturation	98°C	8sec
	Annealing	60°C	20sec
	Extension	72°C	15-30sec/kb
	Final extension	72°C	5min
	End of reaction	4 °C	∞

Table 5. Secondary PCR reaction for RF cloning procedure

2.2.6 Site directed mutagenesis

Site directed mutagenesis was used to mutate a specific sequence of DNA with the desired effect of changing targeted amino acid sequences in the final protein product. For each reaction, complementary primers were designed by Quick Change Primer Design website <https://www.agilent.com/store/primerDesignProgram.jsp> (128). The Quick change II manual protocol was used. Each reaction contained 10-50 ng of DNA, 1 μ L dNTP mix, 5 μ L of 10X Pfu UIltra II reaction buffer, 2 μ L 20 μ M reverse primer, 2 μ L 20 μ M forward primer, 1 μ L Pfu Ultra II polymerase and made up to 50 μ L with distilled water. Then the *DpnI* enzyme was added to the mixture and incubated for 1hour at 37°C followed by transformation into Nova Blue cells. Colonies were selected and the cultures were grown overnight using the appropriate antibiotics, Gen-elute Miniprep kit was used to purified DNA. The DNA sample was sequenced to verify mutation. (129).

	Temperature	Time
	95°C	30 seconds
18 x	95°C	30 seconds
	55°C	1 minute
	68°C	1 minutes /kb
	68°C	3 min
	4 °C	∞

Table 6 Site direct mutagenesis procedure

2.2.7 Transformants analysis (Mini digestion test)

Restriction digestions were performed to analyse transformants. These digestions were performed after purification of the DNA plasmid by mini-prep protocol. These digestions followed the protocol used for the restriction digestions described earlier, but was performed for 1 hour only. The agarose gel electrophoresis was used to view those plasmids with visible and digested inserts (130). These plasmids were sequenced by the Sanger sequencing method in Source Bioscience company.

2.2.8 Gene sub-cloning

All the PCR products were cloned into His-1BKR- pRSF13, His-2GKG- pRSF13, His-MBP- pRSF13 and ubiquitin-pRSF13 vectors using different cloning methods which have been described previously. Each clone used a different oligo primer (Table1).

2.2.9 Protein production

2.2.9.1 Protein small-scale expressions in auto-induction conditions

Small-scale culture protein expressions were carried out in auto-induction conditions for the qualitative comparison due to its ability to efficiently screen the expression and solubility of various clones simultaneously. Small-scale expression cultures were done as follows: 100 mL LB culture with selective antibiotics were incubated at 37°C overnight; next morning, 5 mL of the suspension was transferred into new 100 mL pre-prepared auto-induction media with freshly prepared antibiotics as inoculum then incubated at 37°C at 250 RPM, till the culture reached the mid-point of exponential phase of growth which is ~OD600, then the temperature was reduced to

15°C for a further 24 h to induce protein expression (131). An aliquot was saved as non-induced fraction before decreasing the temperature. The following morning before the cells were harvested an aliquot was kept as an induced fraction. Cells were pelleted for 10 minutes centrifugation at 5000 RPM the cells were harvested and pellets were re-suspended in re-suspension buffer (20 mM Tris, 200 mM sodium chloride, 50 mM Imidazole, 3% glycerol, pH 7.5). Sonication lysed the cells and another centrifugation was carried out as above. An aliquot was kept from the supernatant as soluble fraction and another aliquot was saved from the pellets as the insoluble fraction (132). The filter syringe 0.22 µm and Hi-Trap chelating HP 1 mL Nickel column filtered the soluble fraction. Nickel column was used to purify the target protein by peristaltic pump. The elution samples of all proteins were concentrated to the same volume then all samples were analysed such as aliquots non-induced, induced fractions, the soluble, insoluble and concentrated purified samples loaded on SDS-PAGE gel (133). The samples were prepared by mixing loading buffer (150 mM Tris-HCl, 300 mM DTT, 6% SDS, 0.3% bromophenol blue, and 30% glycerol) with the aliquots samples (134). Then 7 µL was run on SDS –PAGE gel from all aliquots for each protein for comparison.

2.2.9.2 Protein large scale expressions by induction with Isopropyl β-D-1-thiogalactopyranoside (IPTG)

Protein large scale expression experiments commenced with a 100 mL starter culture of LB broth containing selective antibiotic, which was inoculated with cells containing the plasmid of interest. They were incubated in a shaking incubator at 180 RPM, 37°C overnight. Next morning, 5 mL of these starter cultures were utilized to inoculate a fresh 2L LB medium with suitable antibiotics and grown at 37°C at 180 RPM. These cultures were grown until reached the mid-exponential phase of growth

OD600 and then induced by 0.5 M IPTG and the temperature reduced to 18°C overnight. The following morning, cells were pelleted at 4000 RPM for 10 minutes and the supernatant discarded (135). Lysis buffer was used to re-suspend pellets and the sonication technique was used for cell pellet lysis. The sonication was done by Soniprep 150 sonicator machine for 4 min at 25% maximum power. After sonication, samples were centrifuged at 100000 RPM at 4°C for 1 hour (136). After centrifugation, the lysate was filtered by a 0.22 µm filter syringe and the target protein was purified utilizing different purification methods as discussed below.

2.2.10 Protein purification

2.2.10.1 Immobilised metal affinity chromatography

The first purification step was used in this whole study was Immobilised metal affinity chromatography (IMAC). The proteins were purified by IMAC due to the binding properties of metals such as copper, zinc and nickel to the poly-histidine tag present in the protein of interest. The 50 mM lysis buffer contained was used to wash the HisTrap™ High Performance 5 mL Nickel column to eliminate any bound proteins. Then, the protein sample was loaded by peristaltic pump into the column and the column connected to the AKTA machine. The elution buffer containing 500 mM imidazole was used to extract the target protein by concentration gradient (137). SDS-PAGE gel assessed the elution samples.

2.2.10.2 Cation exchange chromatography

Ion exchange chromatography separates charged proteins based on the pH of their environment, as the proteins can have a net negative charge, a net positive charge or no

charge. The choice of pH buffer was used to determine the net charge of the target protein as the pH lower than the isoelectric point of the target protein, this protein carries a net positive charge; therefore, a negatively charged cation exchange resin is selected to bind this protein and vice versa. (138).

In this study, cation exchange was only utilized for ubiquitin protein purification and the isoelectric point of the ubiquitin was calculated by the ProtParam server <https://web.expasy.org/protparam> (139). Before the protein was filled into the column, the column was equilibrated using buffer A (20 mM ammonium acetate, pH 5.1). After the column was equilibrated, the ubiquitin protein was loaded into a HiTrap® SP GE Healthcare 5 mL column by peristaltic pump and connected to the AKTA machine. The ubiquitin protein was eluted by the concentration gradient by elution buffer (0.5 M NaCl, 20 mM ammonium acetate, pH 5.1) (140). The fractions were assessed on SDS PAGE gel.

2.2.10.3 Size exclusion chromatography

Size exclusion chromatography is a technique for the quantification and sizing of proteins and aggregates (141). Proteins of different sizes and shapes were separated on a column matrix depending on their ability to enter and remain within the matrix pores. Superdex gel filtration columns were used for size exclusion chromatography. The used column is determined by the size of the protein. Columns before used were equilibrated with gel filtration buffer (200 mM NaCl, 20 mM Tris pH 7.5, and 1% glycerol). Samples were loaded into a loop and injected into the column using FPLC machines. Gel filtration buffer was used for protein elution by isocratic elution (142). The fractions

were collected and analysed on SDS PAGE gel. The remaining purified protein was stored at -80°C after being immersed in liquid nitrogen (143).

2.2.11 Measurement of protein concentration

The proteins were concentrated in different concentrators based on the protein molecular weight. After the protein was concentrated the protein was measured using a NanoDrop™ 2000/2000c Spectrophotometers machine (Thermo Scientific. MA, US). The Nanodrop was utilized to analyse the purified protein concentration. Amino acids with aromatic rings absorb ultraviolet light with at 280 nm on the other hand, at 200 nm the peptide bonds absorb the ultraviolet light (144). The extinction coefficients were calculated by imputing the primary sequence into Protparam server <https://web.expasy.org/protparam>. The value generated from Nanodrop was divided by the extinction coefficients and the correct concentration was determined.

2.2.12 Protein characterization methods

2.2.12.1 SDS-PAGE

SDS-PAGE gel comprises of resolving and stacking gels, the resolving gel prepared using different acrylamide concentrations. The choice of resolving gel is depending on the protein size to be resolved.

For this study, 15-20% gels were used. The 15 % gels were most commonly used to resolve the purified proteins while for resolving small proteins 18 % gels were used. In all cases, 6% stacking gels were used. The samples of SDS-PAGE were done by mix the protein sample with reducing and non-reducing loading buffer. Before loading samples in the gel wells, the samples were heated at 95°C for 5 minutes. At constant

200 Volt the gels were run till the front band hit the bottom of the gel. Coomassie Brilliant Blue stain G-250 Dye (Thermo Fisher Scientific) were used to stain gels while 50% methanol, 10% glacial acetic were used to de-stained the gels (145).

2.2.12.2 ThermoFluor assay

ThermoFluor is a technique which assesses the thermal stability of the proteins. It measures the melting temperature which is described as the temperature at which the protein unfolds and reveals its hydrophobic core (146). During the assay, the protein of interest is mixed in an aqueous solution with SYPRO Orange dye. SYPRO Orange dye in the aqueous state does not produce a high fluorescence but after the temperature rises the protein denatures and the dye binds to exposed hydrophobic residues at the protein's core and becomes highly fluorescent. The fluorescence can be detected using a real time PCR Agilent Stratagene machine (147). The protein melting temperature is detected by sigmoidal fluorescence curve and its defined as that midpoint of that curve (Figure 15).

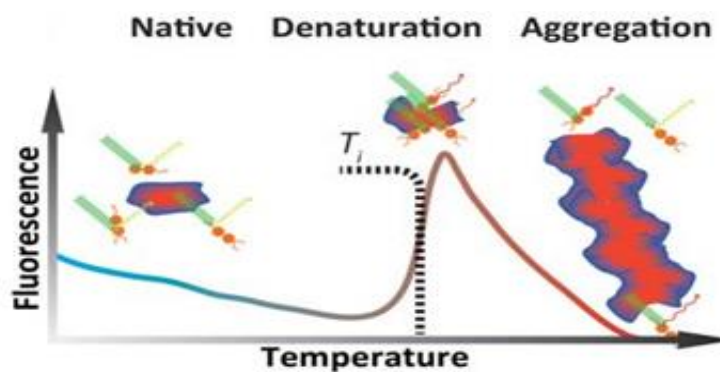


Figure 15. Thermogram revealing different protein states including native, denatured and aggregated states. Figure adapted from (148)

ThermoFluor has been utilized in this study to evaluate the 2GKG tag, 1BKR tag and 1BKR mutant types thermal stability. It was also used to compare the ubiquitin

protein and ubiquitin mutant UbHisS3S4, after the histidine tag was inserted. The experiments were conducted in triplicate for each protein (149).

In capped PCR tubes, 5 µg of the target protein was incubated with 40 µL of the test buffer and 5 µL of the 10 X SYPRO orange dye in a 50 µL total volume. Stratagene Mx3005p RT-PCR machine was used for this assay by using a program with 70 cycles with a 1°C rise from 25 to 95°C.

2.2.12.3 Small-angle X-ray scattering (SAXS) technique

Small angle X-ray scattering (SAXS) technique can be utilized to evaluate protein characteristics such as protein size, shape, degree of aggregation, protein unfolding and folding (150). These can be carried out through Pair-Distribution Function $P(r)$ plot, Guinier plot and Dimensionless Kratky plot (151). The Guinier plot provides geometrical characteristics of proteins including radius of gyration which estimates the average protein size in solution (152). The radius of gyration is predicted from the Guinier plot slope which represents the scattering vector(q) versus logarithm of the scattering intensity(153). The Guinier plot is also useful for the detecting of the protein distribution and protein aggregation (Figure16).

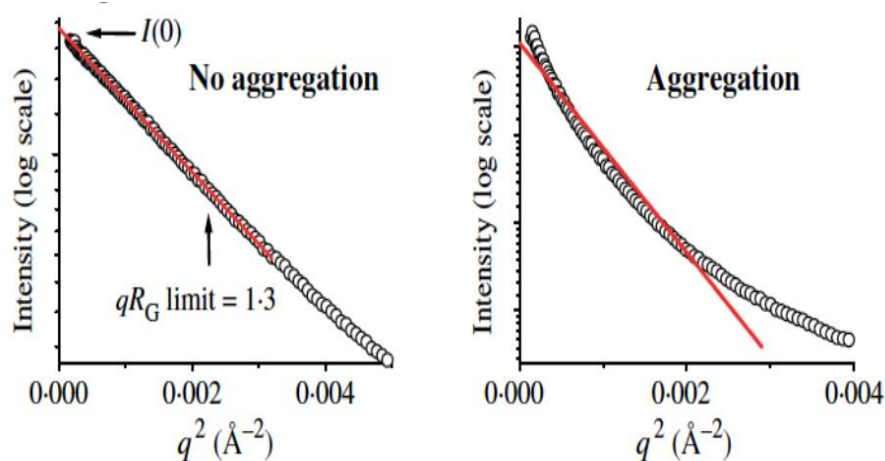


Figure 16. Example of data analysis of the Guinier Plot. Figure adapted from (154).

The normalized kratky plot is informative to verify a protein's globularity and flexibility (155). A bell shaped curve present on a normalized kratky plot is a feature of a well-folded globular protein. However, fully unfolded protein will be characterised as a plateau (156) (Figure 17). Flexible proteins are characterised on a normalized kratky plot by not meeting the q axis (157).

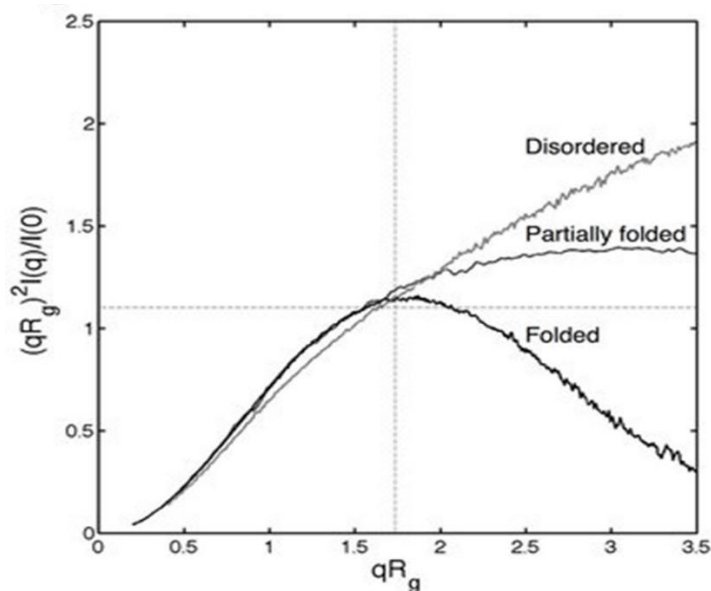


Figure 17. Example of Dimensionless Kratky plots. Figure adapted from (155).

It is also possible to evaluate the aggregation of proteins using the distance distribution function $P(r)$, which is the histogram of all interatomic distances inside the protein and can provide data quality e.g. aggregation (158) (Figure18). It gives specific details about the distances between electrons from the scattering sample and determines the maximum linear dimension (D_{max}) which approximately equal to the maximum protein diameter (159). The D_{max} value contains critical information on the protein shape and compactness of protein(160).

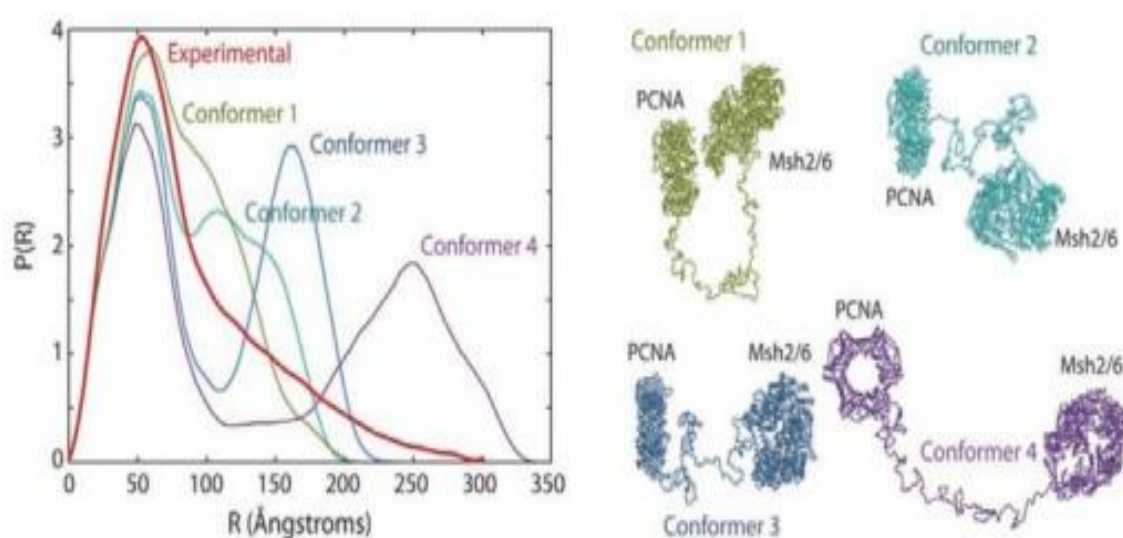


Figure 18.Example of proteins profiles with different structural features in pair distribution function. Figure adapted from (161).

Each protein was dissolved in buffer (20 mM Tris, 200 mM NaCl, and 1% glycerol, pH 7.5) and sent to European Synchrotron Radiation Facility(ESRF), Grenoble. Some samples were performed by different methods such as HPLC and robot. Each protein was measured by three different concentrations. The scattering data from the proteins were picked up by SAXS detector providing a combined q range from $0.005 \text{ \AA}^{-1}$ to $2.0 \text{ \AA}^{-1}$. Scatter Software was used for data analysis such as buffer subtraction, Guinier plot, normalized Kratky plot and Pair-Distribution Function $P(r)$ plot (162).

2.2.13 Other methods

2.2.13.1 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a technique which is used to characterize the thermodynamics of binding protein interactions (163). In one assay, the equilibrium constant (K), enthalpy (ΔH) and stoichiometry (n) can be determined for interactions to be studied (164). Figure 19 shows the general schematic of the titration calorimeter. There are two cells, one a reference cell and the other the sample cell. The reference cell is filled with a buffer, the sample cell is loaded with protein and the stirring syringe is filled with the ligand (165). The stepper motor gradually titrates the ligand into the cell and the stirring maintains homogeneity. The detectors measure the cell temperature continuously and the heater ensures that both cells are kept at a steady temperature (166). If the protein and ligand bind to each other, heat will be produced or absorbed which is depend on the nature of the interaction either endothermic or exothermic (167). This causes the temperature of the cell to change, which is detected and represented by a thermogram. The thermogram represented the injection peaks which is the plot of energy per mole versus ratio of ligand to protein concentrations. ΔH enthalpy, dissociation constant and stoichiometry can be calculated by using the thermogram (168). The difference between upper and lower plateau is the ΔH enthalpy, while the dissociation constant K_d is estimated from the association constant K_a as the relation between K_d and K_a is inversely $K_d = 1/K_a$. The slope of the middle of the sigmoid curve is the K_a while the midpoint of the sigmoid curve is the stoichiometry (169) (Figure 19).

The suggested values for syringe concentrations are 10 X that of the cell in order to obtain successful thermogram peaks and protein saturation (170).

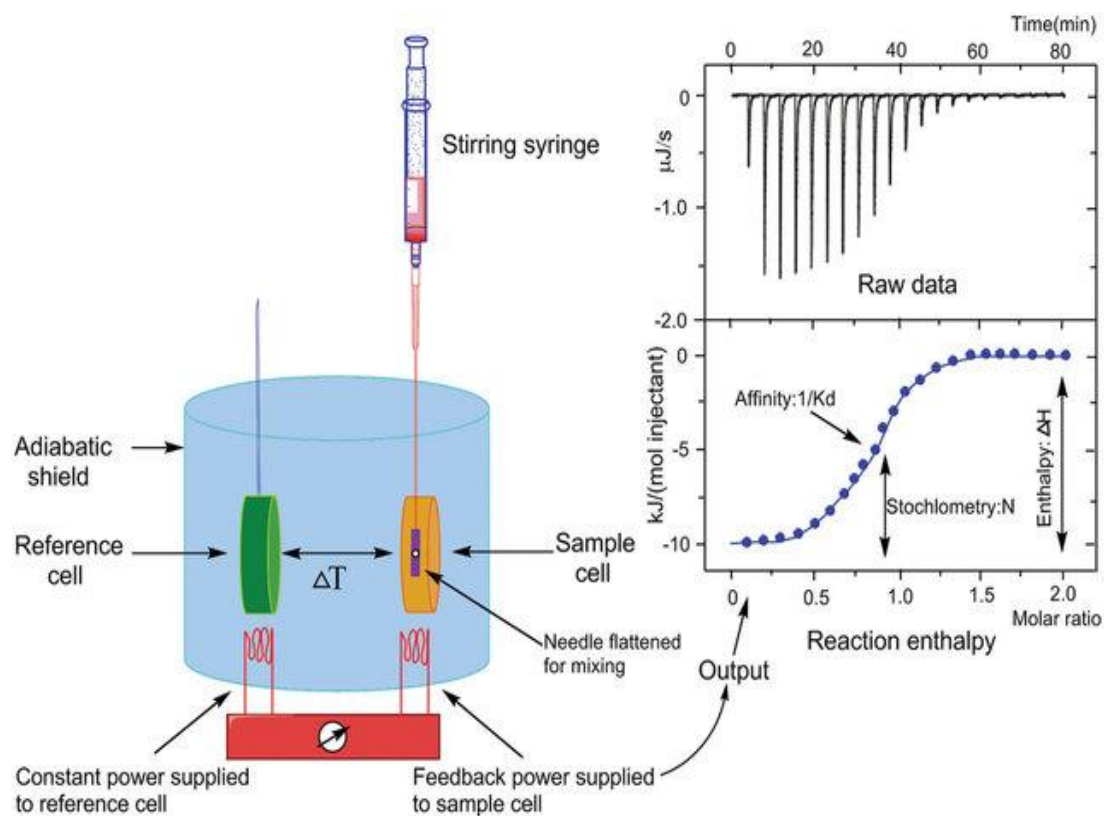


Figure 19. Diagrammatic representative of isothermal titration calorimetry. Figure adapted from (171).

Isothermal titration calorimetry experiments were used in this study to evaluate binding between a modified ubiquitin (with an inserted peptide) and USP11 residues 1-244. The MicroCal VP-ITC machine requires a cell sample of nearly 2 mL and a syringe sample of 0.5 mL but the PEAQ-ITC used here required much lower volumes. Using the same buffer, the USP11 residues 1-244 and the ubiquitin peptide were used on a Superdex 75 gel filtration column. Ubiquitin peptide in the syringe had a concentration of 40 μM and USP11 residues 1-244 at a concentration of 400 μM was filled in the sample cell. The ubiquitin peptide was titrated into the sample cell at 25°C temperature with a stirring speed of 3000 RPM (172).

2.2.13.2 Protein Crystallization

The crystallization of proteins is a technique used to explore protein structures. Protein crystallization has been first observed by German scientists with protein crystals from worms and fish more than 150 years ago (173). It was then developed as an effective tool for purification and chemical purity demonstration in the late 19th century. Protein crystallization was notable for the great achievements with the insulin crystallization in the 1920s and 1930s. (174). Also, Sumner in 1926 found that enzymes could be produced as crystalline proteins (175). In the 1930s, some significant enzymes were purified by crystallization, specially from the pancreas of pigs and cows (176). At the end of the 1930s, protein crystals started to play a valuable new role as a result of the approach of X-ray crystallography introduced to proteins (177). Throughout the 1960s and 1970s, the protein crystals interest grown rapidly as protein crystallography developed and skilled young researchers joined the field (178).

Protein crystallization based on the formation of a super-saturated protein solution (179). Super-saturation can be accomplished by adding precipitating agents and changing different parameters including pH, temperature and ionic strength (180). The hanging drop vapor diffusion technique is the most commonly utilized method for crystallization of proteins (181). Vapour diffusion principle is dependent on the drop consisting of a mixture of protein and precipitant agent at a vapour equilibrium with a liquid reservoir (182). The equilibrium is achieved by diffusing the water vapor from the drop and moving into the reservoir, which increases the concentration of protein and precipitant. After the water moves from the drop, the sample solution in some conditions becomes supersaturated (183) (Figure 20).

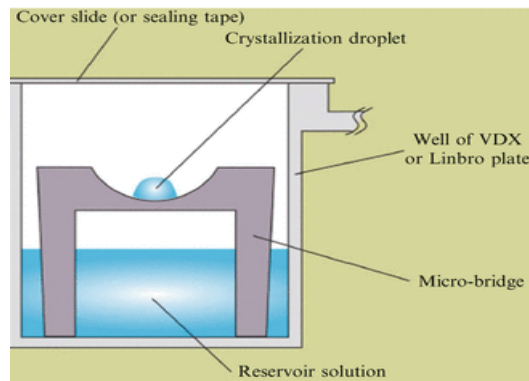


Figure 20.Schematic representation of Vapor diffusion crystallization technique. Figure adapted from (184)

A phase diagram shows the crystallization process (Figure 21). The probability of nucleation and crystal growth are dependent on the concentration of the protein and the precipitant (185). If the concentrations at the initial stage of crystallisation are at the required level, these concentrations can rise until the drop reaches the nucleation phase. When nucleation occurs, the protein is excluded from the solution, which reduces the effective protein concentration (186). It can then reach the metastable zone in which the crystals develop till the protein concentration runs out of this zone (187).

Large screens with several conditions were utilised as different conditions promote the crystallization of different proteins. Conditions include different precipitants such as salts and organic compounds, which are the most widely used precipitants e.g. ammonium sulphate and polyethylene glycol. The conditions may also contain several other compounds including buffers and additives (188).

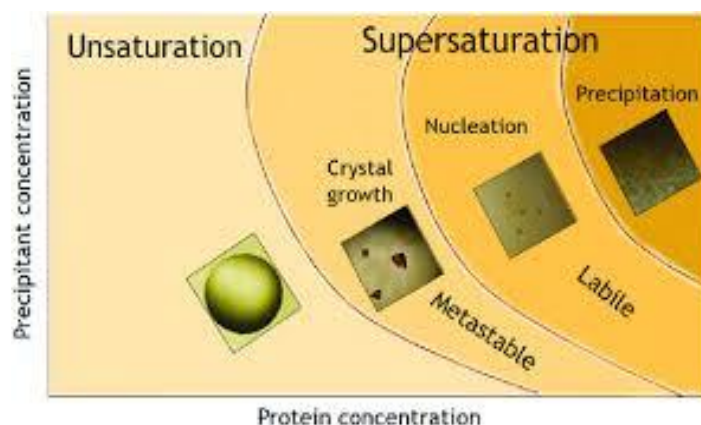


Figure 21. Diagrammatic description of a two-dimensional phase diagram showing the two states of protein solution which are under saturation and super saturation states. Figure adapted from (189)

2.2.13.2.1 Factors affecting crystallization

There are some factors that influence the crystallization of proteins including pH, temperature, precipitants and protein concentrations.

pH scale can affect the protein crystallisation as it can alter the protein molecules charges and solubility (190). Any small change in the pH range can affect the electrostatic interactions between the molecules affecting the molecular packing at the time of crystal formation (191). Several studies have shown that crystal nucleation and growth processes are influenced by the pH of the solution as moving the pH value away from the pI (isoelectric point) increases the region of the nucleation zone (192). The possibility of nucleation increased at 0–2.5 pH units higher than the pI value for acidic proteins and at 0.5–3 pH units lower than the pI value for basic proteins. The possibility of nucleation increased for acidic proteins by 0–2.5 pH units higher than the pI value while for the basic protein by 0.5–3 pH units lower than the pI value (193). Obviously, regulating the pH of a protein solution is a crucial key element in order to ensure efficient protein crystallization (194).

The second factor affecting protein crystallisation is the temperature. The benefits of temperature as a crystallization factor are simplicity of regulation, screening and reversibility (195). The temperature can influence the nucleation and growth of the protein crystals by changing the rate of evaporation of the solvent. In case of high temperature, the evaporation rate will increase and the molecules will move faster so potentially increasing the growth rate of crystals (196). At cooler temperatures, crystals will also grow but evaporation of the water will take more time but the crystal may be significantly better formed/ordered (197). This is not a general rule as the ideal crystallization temperature of every protein is unknown and some proteins can be crystallised at low or high temperature, which is dependent on the solubility of the proteins which is crucial for protein crystallisation (198).

Precipitants are the third factor affect the protein crystallisation. The precipitant's role is to enhance the attractions of protein molecules. This promotes the formation of a crystal packing. Protein precipitants can be categorised into salts, organic compounds and long-chain polymers (199).

The protein concentration also is important factor which affect the protein crystallization as the crystal growth rate can increase in high protein concentrations, however in some cases with low protein concentrations perfect protein crystals were obtained. The low concentrations enhance slower and more controlled growth resulting in better crystals (197).

Finally, whether crystals can be formed or not also depends on the protein. Highly flexible proteins or multi-domain proteins form an ordered array as found in a crystal less easily. The purity of a protein in crystallisation trials is also important (200).

In this study the crystallisation trials were performed in 96 well plates, using multi-channel pipettes. The wells were loaded with precipitant solution. The protein was dispensed by a mosquito ® crystal robot to the wells of a 96-well MRC plate (201). The plate was sealed and stored at 20°C with sealing tape and monitored regularly for crystals.

2.2.14 Cryo-cooling

Cryo-cooling is important to decrease radiation damage to the crystal when exposed to the X-ray beam. Anti-freeze agents such as 15-30% glycerol, ethylene glycol, polyethylene glycols and salts were added to the mother liquor to decrease crystal damage. Specific cryoloops are used to fish the anti-freeze agents-soaked crystal and were immersed in liquid nitrogen before exposure to the X-ray beam (202).

2.2.15 Bioinformatics servers

The amino acid sequences collected from NCBI Blast server were used to calculate the level of sequence identity and similarity of proteins (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (203). The ClustalW server was used for the alignments of the amino acid sequences (https://www.ebi.ac.uk/Tools/psa/emboss_needle) (204). For the sequencing results and amino acid translation Expasy server was used (<https://web.expasy.org/translate>)(205). Expasy protapram server was utilised to calculate the molecular weight and isoelectric point of proteins (<https://web.expasy.org/protparam>)(206). Oligocalc server was used for the estimation the melting temperature of primers (<http://biotools.nubic.northwestern.edu/OligoCalc.html>) (207). For the anticipating the

protein crystallization the xtal pred server was used (<http://xtalpred.godziklab.org/XtalPred-cgi/xtal.pl>)(208) and for protein structure predication Swiss model server was utilized (<https://swissmodel.expasy.org>)(209).

3. Chapter 3: Results and Discussion of preparation and investigation of novel candidate tags 1BKR and 2GKG

This chapter shows the results and discussion of the two candidate tags which were chosen using specific criteria from the Protein Data Bank (PDB) to be used as solubility tags. This chapter shows the results and discussion for each tag in protein expressions, purifications, the assessment of the folding, the relative solubility and stability of each tag and their comparison. The results were used to evaluate if these tags are suitable as solubility tags. Initial studies looked at the expression profiles of the tags in isolation (without fusion) to test the suitability as fusion tags as discussed in a previous study (41).

Proteins from two PDB entries were tested for their suitability as solubility tags; the calponin homology domain from human β -spectrin (PDB code:1BKR) (210) and the receiver domain from *Myxococcus xanthus* social motility protein *FrzS* (PDB code:2GKG) (80). These proteins will be referred to by their PDB codes throughout.

The experiments were performed in small-scale (100 mL culture) and large-scale cultures (2 litres cultures). Experiments at the small culture scale were designed to gain a basic understanding if the tags could be expressed, solubilised, purified and for qualitative comparison between the tags (211). The small-scale culture experiments used auto-induction conditions, as this method allows for efficient screening and comparison of the expression and solubility of various clones simultaneously (212).

Large-scale experiments used bacterial cultures grown at a 2 L scale and protein expression was induced upon the addition of Isopropyl β -D-1-thiogalactopyranoside (IPTG) for quantitative comparisons and to facilitate production of high-yields of soluble proteins for subsequent crystallisation trials (213). The large-scale trials assessed the tags performance in terms of expression, solubility, stability and crystallization allowing the fusion proteins to be used for structural studies (214).

A thermal stability assessment was done using Thermoflour assay for both tags that provides information on how the tags can be used for the crystallization of fusion proteins as the highest stability of the fusion tag enhance the formation of crystals of the fusion protein (149).

3.1 Cloning of 1BKR and 2GKG

The 1BKR and 2GKG genes were cloned separately into a pRSF-13 DNA vector after a 6x Histidine tag at the N terminus. The 1BKR tag has a molecular weight 13 kDa and is composed of 131 amino acids, whereas, the 2GKG tag has a molecular weight of 14 kDa and is composed of 140 amino acids. This cloning used the restriction enzyme method as stated in the material and methods chapter.

3.2 Expression, solubility and purification test of 1BKR and 2GKG tags in small-scale auto induction cultures

Small-scale expression cultures used auto-induction conditions for the qualitative comparison of the two tags in terms of expression, solubility and purification (215). SDS PAGE gel investigation of the expressed proteins revealed an extra protein band in the post-induction fraction, no band was observed before induction, which indicated

that both proteins had been expressed (Figure 22 A, B). The observed bands were of the expected size of approximately 13 kDa for 1BKR and 14 kDa for 2GKG, when compared with marker bands.

A solubility test was performed using lysis buffer (20 mM Tris, 50 mM imidazole 200 mM NaCl and 3 % glycerol, pH 7.5) on a cell pellet from an auto induced culture. Analysis revealed an intense band which was analogous to the band expected for the soluble fractions of 1BKR and 2GKG tags (Figure 22 C,D) which indicated that both tags had been expressed in soluble form (216). The lysates were then purified using Nickel column affinity chromatography. The purified proteins were concentrated and signs of aggregation were observed for the 1BKR tag, which affected the purification yield. Analysis of the proteins on SDS PAGE gel revealed that the band analogous to that expected for 2GKG was more intense than that observed for the 1BKR tag. This suggests that the 2GKG tag is more soluble and expresses better than the 1BKR tag (Figure 22 E, F). The total yield from a 100 mL of culture for the 1BKR tag was 0.16 mg of soluble protein and for 2GKG tag it was 3.17 mg of soluble protein.

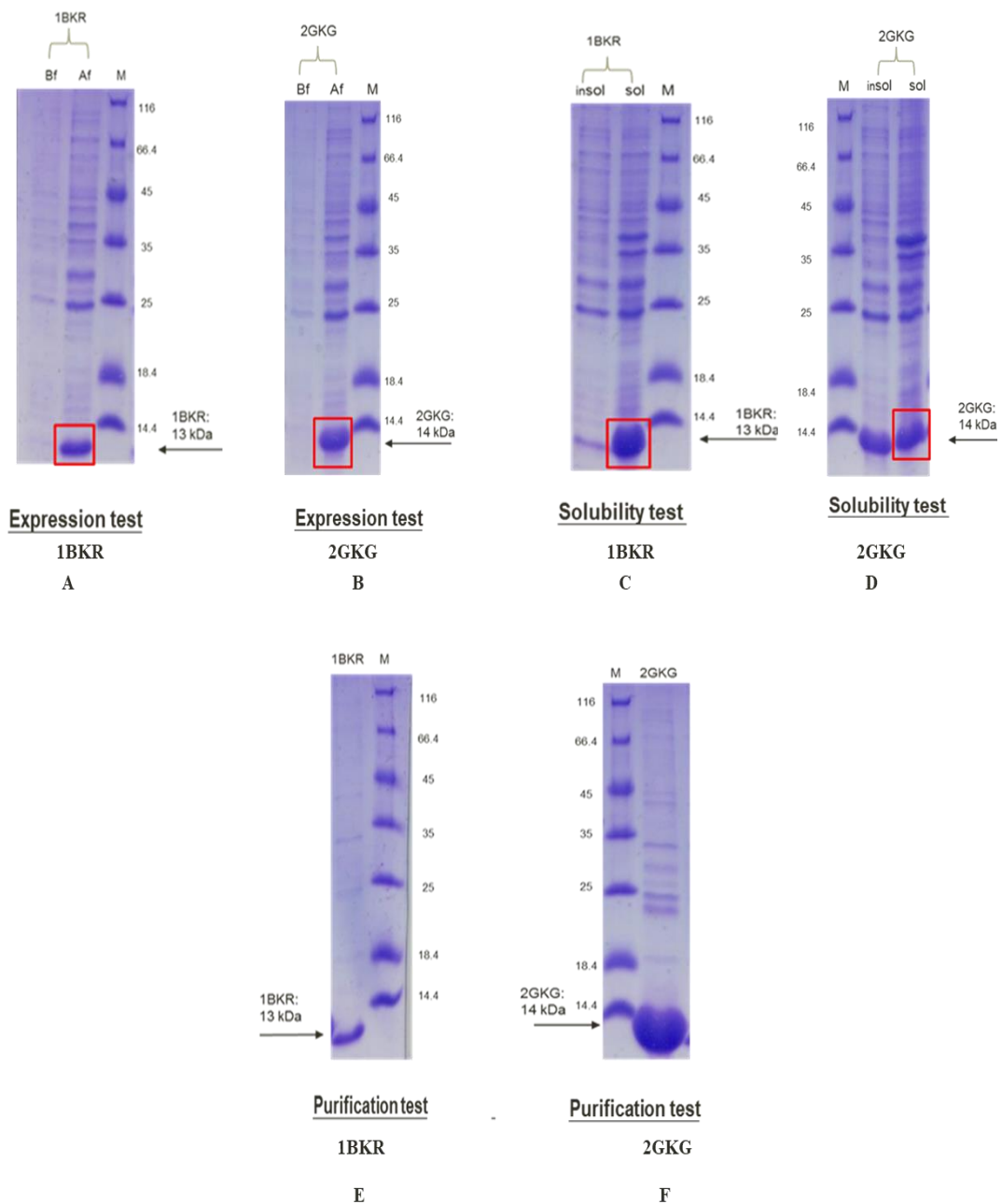


Figure 22. A: Analysis of the un-induced and induced fractions from protein expression experiment of 1BKR tag on SDS-PAGE gel. B: un-induced and induced samples of 2GKG protein were run on a 15 % SDS-PAGE gel C: SDS-PAGE gel of the 1BKR tag solubility test. D: The SDS-PAGE gel of the 2GKG tag solubility revealed protein bands in both the insoluble and soluble fractions E: The SDS-PAGE gel of 1BKR tag purification indicated that the 1BKR tag had been successfully purified. F: The SDS-PAGE gel of 2GKG tag purification revealed a band corresponding to that expected for 2GKG tag.

3.3 Expression, solubility and purification of 1BKR and 2GKG tags in large-scale cultures

For more quantitative comparison between the two tags, the large-scale culture experiments were done using two liters of cultures. The results showed that both proteins expressed well and that after purifications a single peak on nickel-affinity purification (not shown) and a monodisperse peak on gel filtration (Figure 23 A,C) were observed, respectively. After the two purification steps, both proteins 1BKR and 2GKG were very pure, as indicated on SDS PAGE gel analysis (Figure 23 B,D). The molecular masses of 1BKR and 2GKG are 13 and 14.0 kDa, respectively and their elution volumes from the HiLoad 16/60 Superdex 75 gel filtration column were 92 and 90 mL, respectively. These elution volumes correlated with proteins of molecular weight of about 13 kDa and 14 kDa, according to a standard curve (Figure 24). The total yield of the 2GKG protein was 35 mg/L culture while 1BKR protein final yield was 1.8 mg/L culture, which suggested that 2GKG was more soluble than 1BKR. During the concentration step, 2GKG was concentrated without any signs of aggregation. However, when 1BKR reached a concentration of 15 mg/mL aggregation in the form of precipitation was observed. This led me to further investigate to improve the solubility of 1BKR. The investigation was performed on the 1BKR structure using the Pymol software. This revealed that the 1BKR structure has a hydrophobic region with exposed hydrophobic leucine 17, tryptophan 18 and methionine 21 residues (Figure 25). Site directed mutagenesis was performed on a leucine residue (L17) in this region, which was mutated to two different hydrophilic residues, which were serine and aspartate (L17S and L17D) (217). The reason for the choosing leucine 17, was due to its position on the surface of the protein and being not involved in the core protein structure (Figure 25).

This suggested that any changes in leucine 17 would not affect protein stability, but would reduce the hydrophobicity on the protein surface potentially aiding solubility (218).

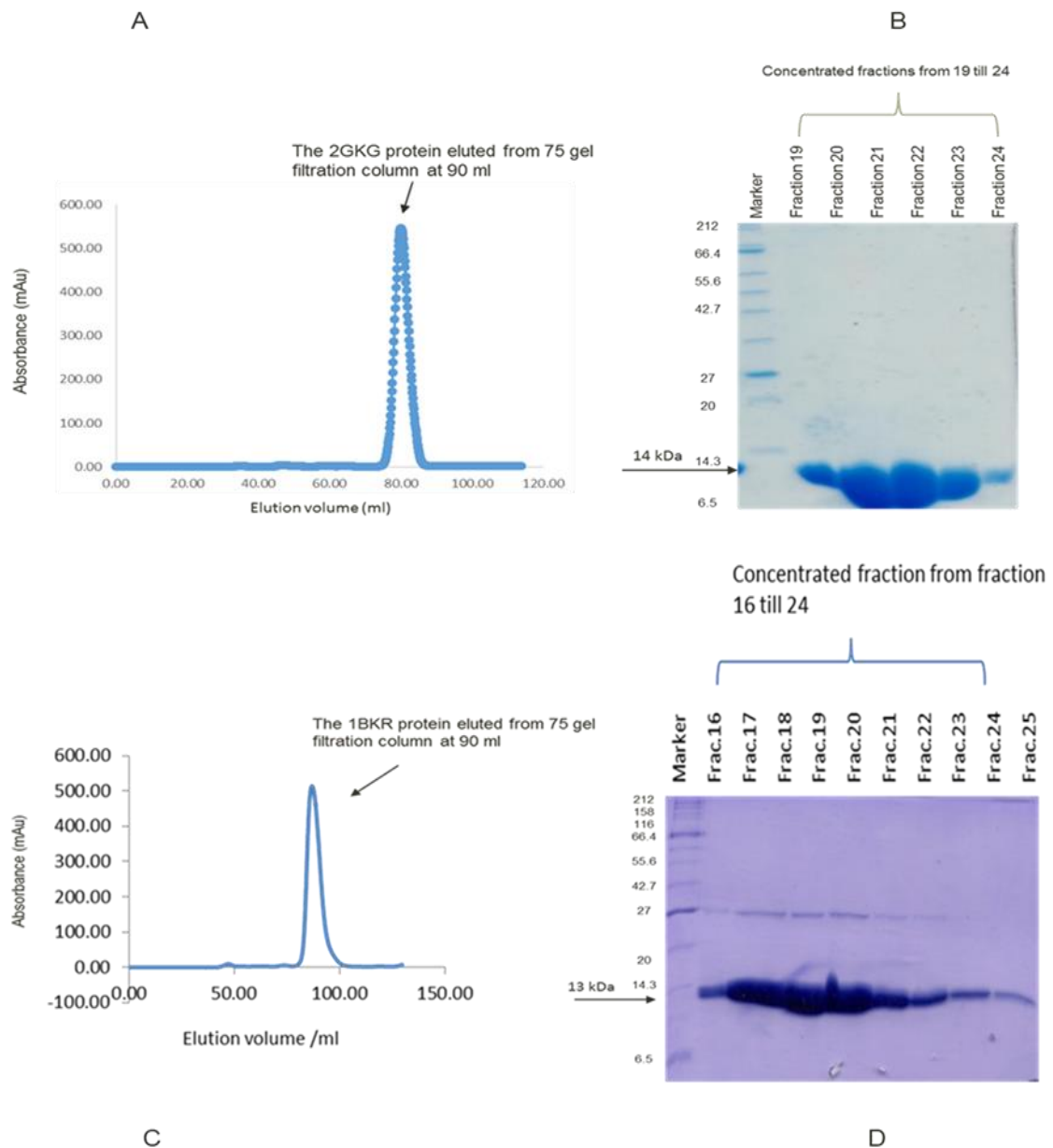


Figure 23. A: Chromatogram of the gel filtration purification of 2GKG. The Chromatogram showed a monodisperse peak containing the protein target. B: SDS-PAGE gel investigation of the peak fractions revealed bands, which corresponded to the expected size of the 2GKG protein (14 kDa). C: Chromatogram of the 1BKR protein after gel filtration purification. D: SDS-PAGE gels showed an intense band below the marker band of 14.3 kDa which corresponds to the 1BKR protein.

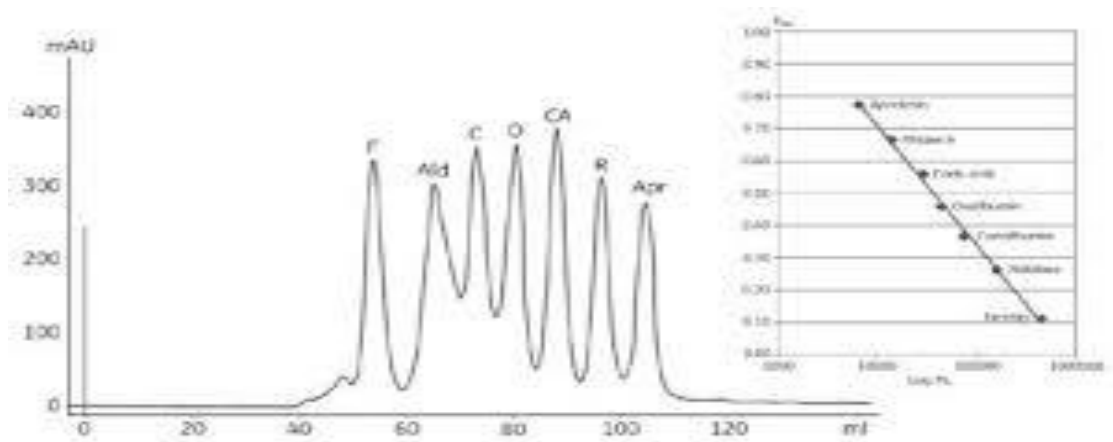


Figure 24. Standard curve for Superdex 75(16 /60) column showing the elution volumes of the different molecular weight of proteins used as standards.

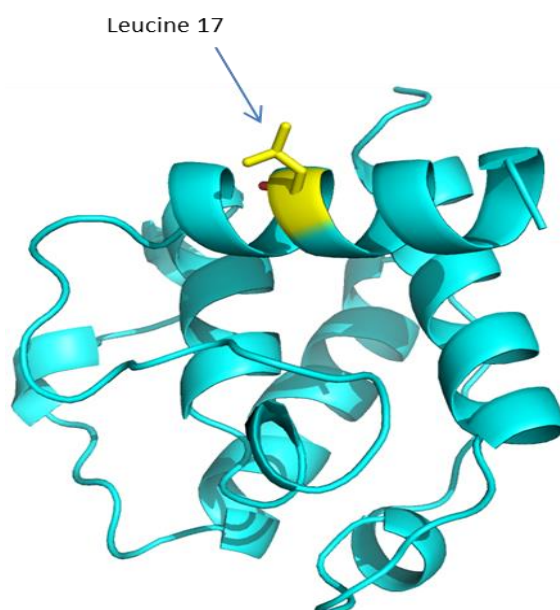


Figure 25. Cartoon diagram of the crystal structure of 1BKR , the leucine residue number 17 highlighted in yellow colour (made in Pymol software).

3.4 Site directed mutagenesis of 1BKR tag

The 1BKRL17S and 1BKRL17D were constructed from the 1BKR gene using site-directed mutagenesis (material and methods). After mutagenesis, sequencing revealed that the new sequence of 1BKRL17S and 1BKRL17D were correct in terms of replacing a leucine residue with a serine residue and aspartate residue.

3.4.1 Expression and purification test of 1BKRL17S and 1BKRL17D in large-scale cultures

In a two litre culture, both proteins expressed well as investigated by SDS PAGE gel (not shown). The purification of the 1BKRL17S and 1BKRL17D used Nickel affinity chromatography initially and then a subsequent gel filtration chromatography as a final purification step. The gel filtration results showed a monodisperse peak for both proteins which eluted at 90 mL. Compared with a standard curve, the elution volume of both 1BKRL17S and 1BKRL17D corresponded to 13 kDa (Figure 26 A,C) (219). The SDS -PAGE gel showed a molecular weight of the proteins of around 13 kDa as predicted (Figure 26 B, D). The SDS-PAGE gel suggested that the proteins were pure as no presence of other protein bands was observed. The yield of 1BKRL17S was 0.5 mg/L culture while 1BKR L17D was 3.2 mg/L. These results showed that the 1BKRL17S yield was lower than that observed for the 1BKR wild type protein while 1BKRL17D had a higher yield than 1BKRL17S and 1BKR wild type.

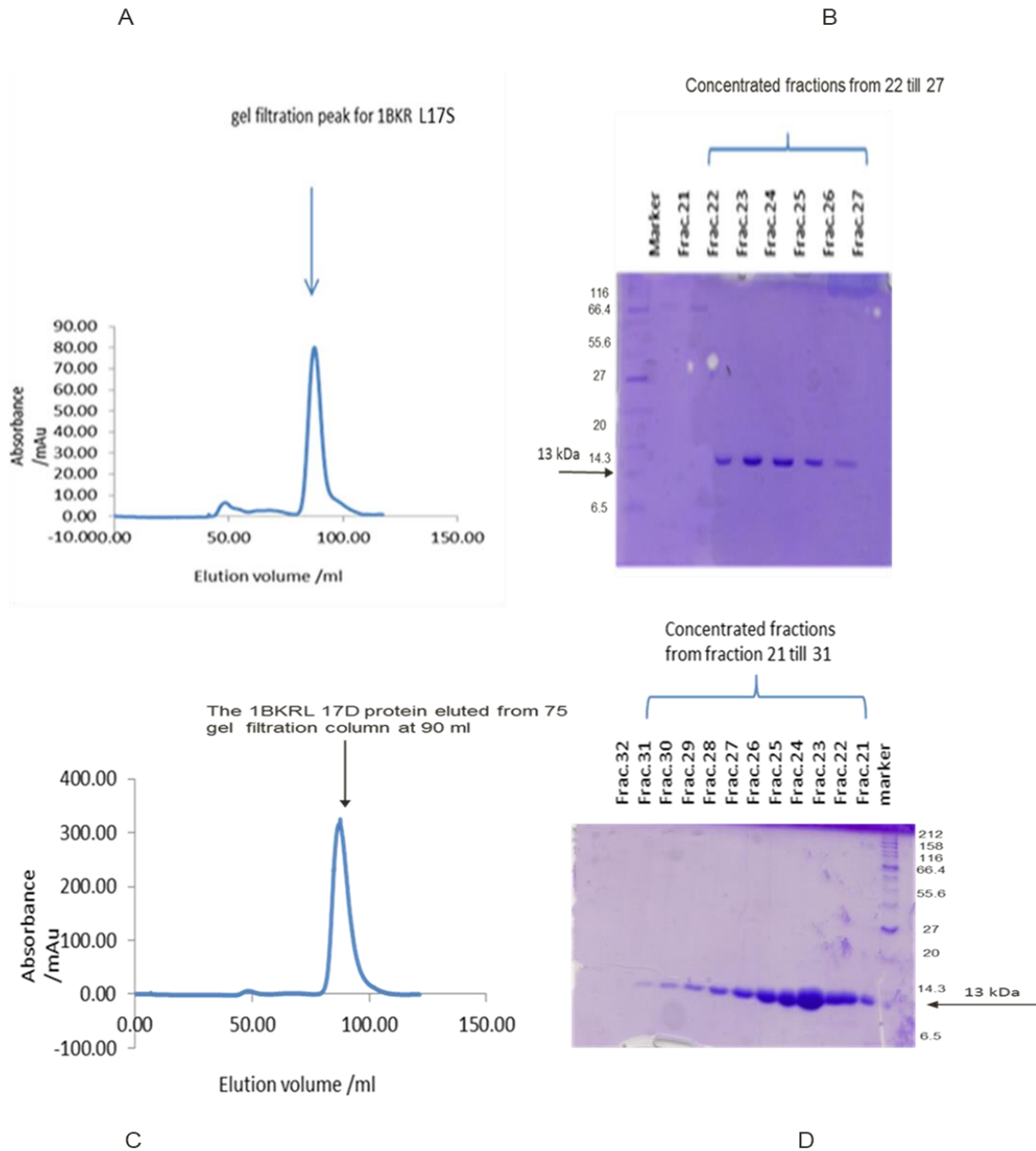


Figure 26. A: Gel filtration chromatogram of 1BKRL17S revealed one peak at 90 ml as elution volume corresponding to 13 kDa based on the standard curve. B: SDS-PAGE gel analysis revealed that the protein was approximately 13 kDa that corresponded to 1BKRL17S C: Gel filtration chromatogram of 1BKRL17D showed a high peak containing the target protein. D: SDS-PAGE gel result showed bands that corresponded to 1BKRL17D (13 kDa).

3.4.2 Stability assay (Thermofluor assay) for 1BKR, 1BKRL17S, 1BKRL17D and 2GKG proteins

Thermofluor assays were performed to analyse the thermal stability of proteins and identify which protein was more stable as indicated by a higher protein melting temperature. Previous research has shown that increasing a protein's stability determined by Thermofluor correlates with an increased crystallisation success of the protein (220). I used this approach to assess the protein stability of the 2GKG tag and 1BKR tag and its mutants and to determine the suitability of 1BKR wild type, 1BKRL17D and 1BKRL17S for subsequent crystallization trials. The mean melting temperature of the 1BKR wild type was 47.1°C, which is very close to the 1BKRL17D and 1BKRL17S mutants which were 46°C and 48.6°C, respectively. There was no significant difference in average melting temperature ($p = 0.224$) which suggests that there was no significant difference between the thermal stability of the three proteins using this method (Figure 28). Stability assays revealed that insertion of either serine or aspartate at residue 17 had no effect on 1BKR protein stability when compared with the wild-type protein. This data revealed that all these proteins were suitable for subsequent crystallization trials based on a previous study that reported that proteins with melting temperature higher than 45°C can be crystallized (147).

The mean melting temperature of 2GKG was 45.2°C and there was a significant difference between the 2GKG and 1BKR wild type ($p = 0.0005$) proteins, potentially suggesting that the 1BKR wild type protein had a higher tendency to form crystals than the 2GKG protein based on melting temperatures alone.

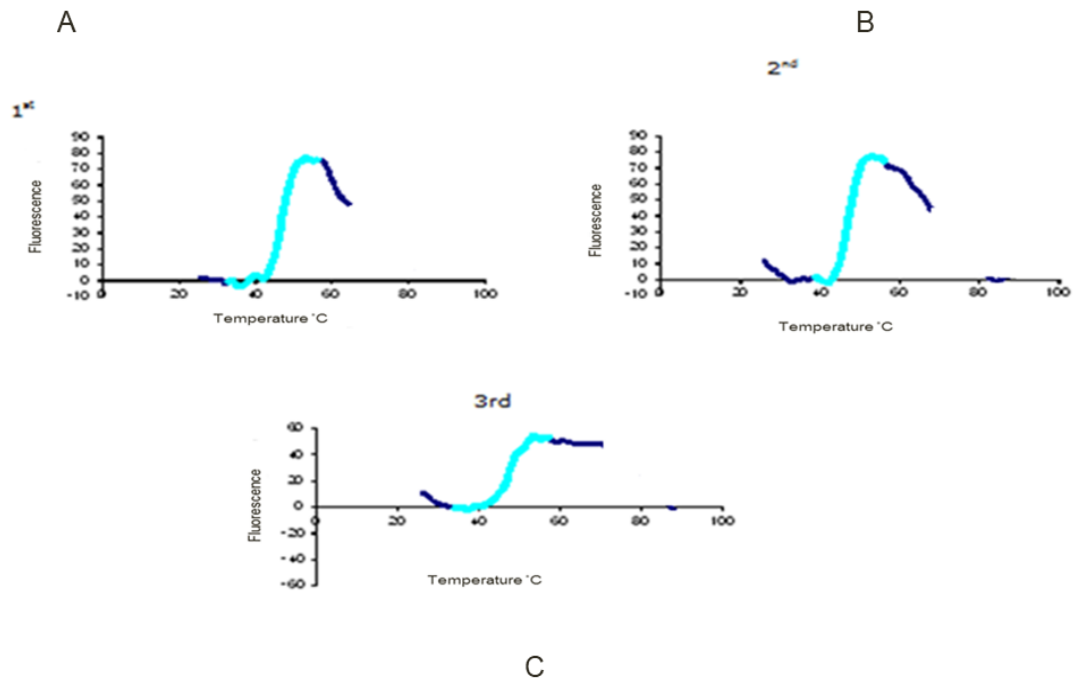


Figure 27. Sample of raw data from the ThermoFlour assays showing the measurement of the T_m which is the temperature where the protein unfolds revealing its hydrophobic core which is defined as the midpoint of the curve. The melting temperature represented by the turquoise colour while the blue colour after turquoise colour refers to protein aggregation. A: the first measurement B: second measurement, C: third measurement.

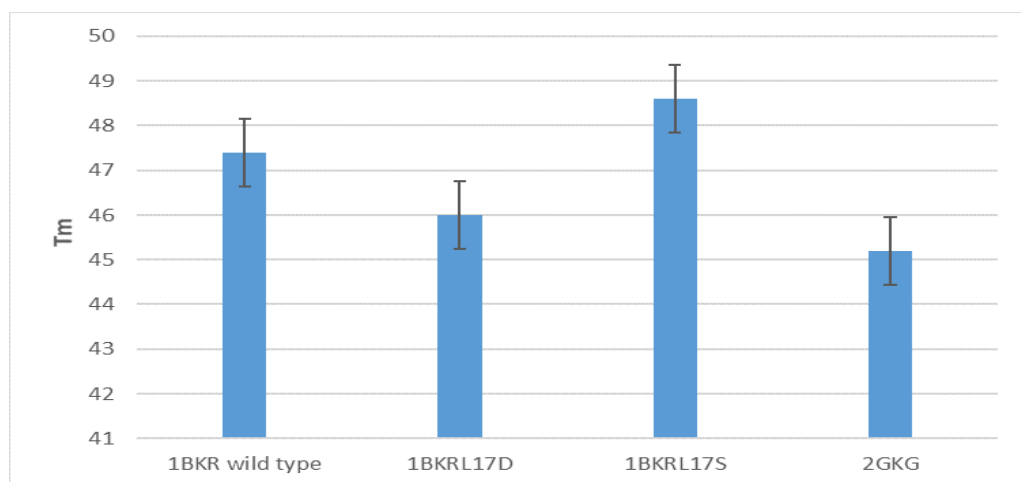


Figure 28. Bar graph representing the melting temperature measurements from the ThermoFluor experiments of 1BKR wild type, 1BKRL17D, 1BKRL17S and 2GKG proteins.

3.5 Crystallization trials of 1BKRwildtype,1BKRL17D, 1BKRL17S and 2GKG for crystallization

The aim of the crystallisation trials was to investigate the suitability of 1BKR wild type, 1BKRL17D, 1BKRL17S and 2GKG proteins for crystallization. An assessment of the potential of these proteins in terms of crystallization would be important criteria for determining if the fusion tag would be amenable to act as carrier protein for crystallizing model proteins. Several screens were conducted with each protein such as PACT premier, Morpheus, PEG II Suite, Structure, MIDAS, PGA, Stura FootPrint Eco, and JCSG plus. After incubation of one-month at 20°C, crystals formed in two conditions in the 1BKR wild-type trials when using a Stura FootPrint Eco screen (Figure 29 A, B). However, no crystals were obtained with 1BKRL17D and 1BKRL17S which suggests that 1BKR wild type can form crystals with fusion proteins better than 1BKRL17D and 1BKRL17S proteins. For 2GKG tag crystallization trials, the trials replicated those performed with the 1BKR tag with the same screens but using a higher concentration as the 2GKG protein purified in higher concentration than 1BKR tag, one crystal was formed when using a JCSG plus screen (Figure 29 C).

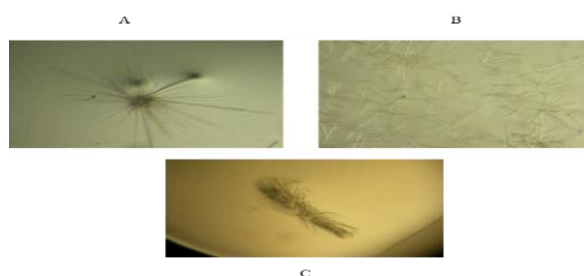


Figure 29. (A) Crystal of 1BKR wild protein in condition: conc. Salt: 1.5M sodium citrate tribasic dehydrate, conc. Buffer: 0.01 M sodium borate, pH 8.5. (B) Crystals of 1BKR wild protein in condition: conc. Salt: 1.2M sodium citrate tribasic dehydrate, conc. Buffer: 0.01 M sodium borate, pH 8.5. (C) Crystal from 2GKG protein on JCSG+ screen on the condition is 0.02 M Calcium chloride dehydrate, 0.1 M sodium acetate, 30 % v/v MPD pH 4.6

3.6 Discussion of 1BKR and 2GKG protein expression and crystallisation results

This chapter investigated the preparation and expression of 1BKR and 2GKG, which were selected from the PDB as potential candidate solubility tags. Both tags were cloned into the pRSF13 vector at the N terminus position after the His tag (affinity tag), which was used for affinity purification for both proteins. The histidine affinity tag was chosen as it has a relatively small size, which rarely affects fused protein function and promotes the expression and purification of fused proteins more than that observed for other affinity tags (48). The N-terminal position was selected for the tags as it gets translated before the fusion partner which may promote better folding and stability of the target proteins in comparison to a C terminal tag (221).

Both proteins were expressed and purified from small and large-scale cultures. The small- scale experiments being important as valuable indicator for anticipated protein production yield in large-scale expressions. The results from the small-scale trials revealed that 1BKR and 2GKG were expressed successfully as soluble proteins. After purification and concentration, a sign of aggregation was observed with the 1BKR tag, which affected the final yield when compared with 2GKG tag. This data suggests that 2GKG after purification and concentration was more stable and soluble than 1BKR (6).

Data derived from large-scale cultures allowed for a quantitative comparison between the two novel tags. The conclusion that the 2GKG tag was more stable and

soluble than 1BKR was observed to aggregate when concentrated above 15 mg/mL, while this effect was not observed for the 2GKG tag.

To overcome this aggregation issue during concentrations of the 1BKR tag, residue present in a hydrophobic region which is leucine number 17 was mutated to serine and aspartate. Previous research has highlighted the importance of mutation to hydrophilic serine and aspartate amino acids for improving protein solubility (222) (223). The presence of the mutations had no effect on stability and protein folding as the Thermoflour assay and gel filtration results indicated, however, the highest yield obtained was 3.18 mg/L for 1BKRL17D, 1.8 mg/L for 1BKR wild-type and 0.5 mg/L for 1BKRL17S (Table 7). This data suggests that the mutation of leucine to aspartate (1BKRL17D) improved soluble yield and no aggregation was observed. Crystallisation trials revealed that the 1BKRL17D mutant protein failed to form crystals. On the other hand, 1BKR wild type formed crystals in two conditions which suggested that the mutations had affected the proteins ability to form crystals. This could indicate that the minor crystal contacts that Leu17 is part of in the wild-type crystal structure may promote crystallisation. This indicated that the 1BKR wild type had an advantage to be used as a fusion tag for crystallisation studies as carrier driven crystallisation using fusion tags for crystallisation can sometimes help (214). The 2GKG protein formed crystals in one condition which suggest that both tags can be used to crystalize fusion proteins which have been designed with the fixed arm linker, presence of the linker has facilitated crystallization of many fusion proteins when fused with a MBP tag (224).

Thermostability comparisons revealed that the 1BKR wild type was significantly more thermally stable than 2GKG with both proteins having a melting temperature

which was higher than 45°C. The 1BKR wild type protein average melting temperature was 47.1°C and 2GKG was 45.2°C. This result suggests that 1BKR can form crystals potentially more easily than 2GKG, as it has a higher melting temperature, which can promote the crystallizations of proteins. This could be one explanation for the ability of 1BKR to form crystals in different conditions more readily than the 2GKG protein (220).

A final conclusion is that 1BKR wild type and 2GKG proteins in isolation were expressed well with a high yield and formed crystals which may suggest that the presence of these tags could help protein folding and solubility of a fusion protein. The assumption is based on results observed for Skp and T7PK tags which were expressed and purified in high yield as single tags (without any fusion) and when they were fused to target proteins they improved their expression, solubility and purification (41). The following chapters discuss the performance of 1BKR and 2GKG tags when fused to different chosen model proteins with a fixed arm linker in terms of expression, solubility, purification and crystallization of the fusion proteins.

Construct	Yield (mg/L)
1BKR wild type	1.8
1BKRL17S	0.5
1BKRL17D	3.18

Table 7. Comparison of the yields of 1BKR wild type, 1BKR L17S and 1BKR L17D constructs after gel filtration purification using Superdex 75 HiLoad 16/60 column.

4. Chapter 4: Results and discussion of the fusion of 1BKR and 2GKG tags to model proteins UCHL1 and UCHL3

This chapter focuses on the investigation and comparison of the candidate protein tags i.e. 1BKR and 2GKG, with two model globular proteins of known structure and compares the performance of these novel tags with the performance of these proteins when tagged with MBP, which is a commonly used tag (51). The investigation and comparison will be in terms of expression, solubility and purification in small-scale bacterial cultures in auto-induction conditions for qualitative comparison. As before for the tags in isolation, large-scale 2 litres bacterial cultures for quantitative comparisons to produce high-yields of soluble proteins for subsequent crystallisation trials for each model protein were also performed.

The model proteins chosen were ubiquitin carboxyl-terminal hydrolase isozyme L1 (UCHL1) (PDB code 2ETL) and ubiquitin carboxyl-terminal hydrolase isozyme L3 (UCHL3) (PDB code 1UCH). These model proteins were chosen as their structures are known and relatively globular. In addition, these proteins are of interest as drug targets and increased yield and potentially novel crystal forms are desirable. These experiments were carried out to study the effect of these novel tags on the expression, solubility, purification and crystallization of these two model proteins. The potential crystallization of these two model proteins in different space groups (crystal packing) may be better suited for compound soaking (225).

The UCHL1 enzyme contains 223 amino acids and has a molecular weight of 24.8 kDa. The isoelectric point of UCHL1 is 5.33. The UCHL3 consist of 230 amino acids with a molecular weight of 26.1 kDa and an isoelectric point of 4.84.

All the proteins were purified by Nickel affinity chromatography as a first step purification followed by gel filtration chromatography. Nickel affinity chromatography was selected because the histidine affinity tag rarely impacts fused protein function letting it to be a favourable than other affinity tags (226). The gel filtration was the second step of the purification to ensure that the protein is stable, not aggregated and can be used for crystallization trials. (227). This is can achieved by monitoring the elution volume of each protein as each protein has expected elution volume which depends on the size and shape of the protein and can be compared to a standard curve of each column. If the protein elutes at the void volume peak this means that the protein forms soluble aggregates and is not suitable for further experiments (228). The gel filtration step was done only in the large scale culture as we need to assess the protein stability and purification to be used for crystallization trials and for quantitative comparison while the small scale trials were only done for qualitative comparison.

4.1 Results of investigation of UCHL1-tag fusions

4.1.1 Cloning of 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1 constructs

Three UCHL1 constructs were generated, namely 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1. The 1BKR-UCHL1 was cloned into a pRSF-13 vector after histidine tag at the N terminus.

The second construct was 2GKG-UCHL1, which was cloned the same way as 1BKR-UCHL1 but with a His-2GKG tag at the N terminus. The UCHL1 gene was successfully inserted into the pRSF13 vector as confirmed by the size of DNA fragments following a restriction digest of the vector. Sequencing the inserts confirmed that the ligation and transformations had been successful.

The construction of MBP-UCHL1 used a restriction enzyme free cloning protocol as stated in the material and methods chapter where the 2GKG-UCHL1 construct was used to create the MBP-UCHL1 construct by replacing the 2GKG tag with a MBP tag. Sequence analysis proved that MBP-UCHL1 had been cloned correctly.

4.1.2 Expression, solubility and small-scale purification of 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1

These experiments were performed to make a qualitative comparison between 1BKR, 2GKG and MBP tags when fused with the UCHL1 protein by auto induction protocol, which is more accurate for comparison in terms of expression, solubility and purification (215). Thus, this experiment compared three different constructs 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1 at a small scale on the expression, solubility and purification of model proteins

Analysis of induced samples using SDS-PAGE revealed that all the proteins were expressed and an extra band post-induction when compared with samples taken from non-induced cultures was observed (Figure 30 A). All post-induction bands for each protein were of the correct size, which correlated to the expected molecular weight which was 66.2 kDa for MBP-UCHL1, 40.4 kDa for 2GKG-UCHL1 and 39.6 kDa for 1BKR-UCHL1. Analysis of a SDS-PAGE gel revealed that the protein band from the soluble fractions of MBP-UCHL1 and 2GKG-UCHL1 were more intense than 1BKR-UCHL1 (Figure 30 B). The cell lysates were subsequently purified using a Nickel column. Analysis of the purified samples on SDS-PAGE revealed a more intense band for MBP tag than that observed for 1BKR and 2GKG (Figure 30 C). Analysis also revealed that there was more purified 2GKG tag when compared with purified 1BKR tag which suggests that 2GKG is more soluble than 1BKR tag, as observed in the results discussed in Chapter 3.

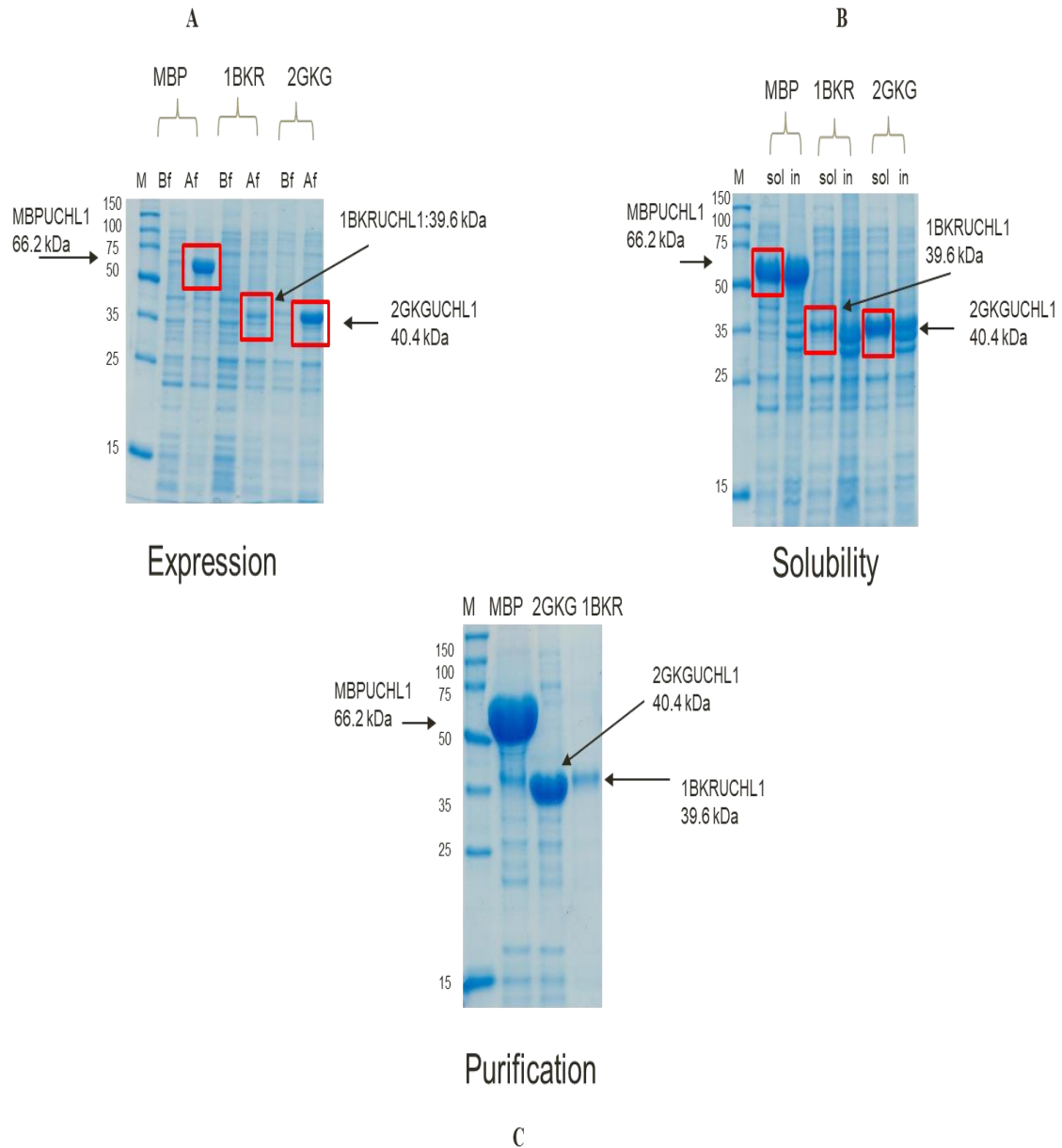


Figure 30.A: The SDS-PAGE gel analysis of expression levels of 1BKR-UCHL1, 2GKG- UCHL1 and MBP- UCHL1. SDS-PAGE indicated all the proteins were expressed as indicated by the presence of bands after induction where no bands were present before induction. **B:** SDS-PAGE gel analysis of the soluble fractions of 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1, which revealed bands in the soluble fractions. Comparing soluble fractions revealed that MBP- UCHL1 and 2GKG-UCHL1 were more intense than that observed for 1BKR-UCHL1. **C:** SDS-PAGE analysis of samples after purification of 1BKR-UCHL1, 2GKG- UCHL1 and MBP-UCHL1 revealed that the band analogous to the band of the MBP tag was more intense than that observed for the other tags.

4.1.3 Large-scale expression and purification of UCHL1 protein-solubility tag fusions

The aim of this experiment was to allow quantitative comparison between 1BKR-UCHL1, 2GKG- UCHL1 and MBP-UCHL1 proteins in large scale and production of high protein yield which could then be used for crystallization trials (229). Three UCHL1 protein solubility-tagged constructs were expressed in two litres volumes and were induced upon addition of IPTG as mentioned before. The proteins were purified using Nickel affinity chromatography as a first step of purification followed by gel filtration chromatography.

4.1.3.1 1BKR-UCHL1 protein purification

Purification of 1BKR-UCHL1 protein using Nickel affinity chromatography revealed a broad single peak indicating high protein yield was obtained (Figure 31 A). The 1BKR-UCHL1 protein eluted using elution buffer 20 mM Tris, 200 mM sodium chloride, 500 mM imidazole, 3% glycerol at pH 7.5. Analysis of the peak fractions on the SDS-PAGE gel, revealed that a protein bound to the nickel column as no band was present in the flow through fraction. Analysis of samples on SDS page revealed the presence of a prominent band between the 30 and 40 kDa protein marker bands, which was analogous to the protein target. The gel also revealed presence of contaminating proteins (Figure 31 B). The gel filtration chromatography was performed on the purified fractions as a secondary purification step to remove the contamination and to analyse the protein stability (230). Gel filtration purification of 1BKR-UCHL1 revealed two peaks. The first peak was void volume (Figure 31 C) and the second peak contained the protein of interest, this protein eluted at elution volume of 65 mL, from the gel filtration

column (Superdex75 (16/60)). This volume correlates to the standard curve prediction of a protein with a molecular weight of about 39.6 kDa. This result also indicated that no aggregation of the target protein as the protein eluted in the correct elution volume not in the void volume peak. The SDS-PAGE gel also confirmed that the product contained some degradation proteins (Figure 31 D). The total pure protein after concentration was 12.45 mg/L culture.

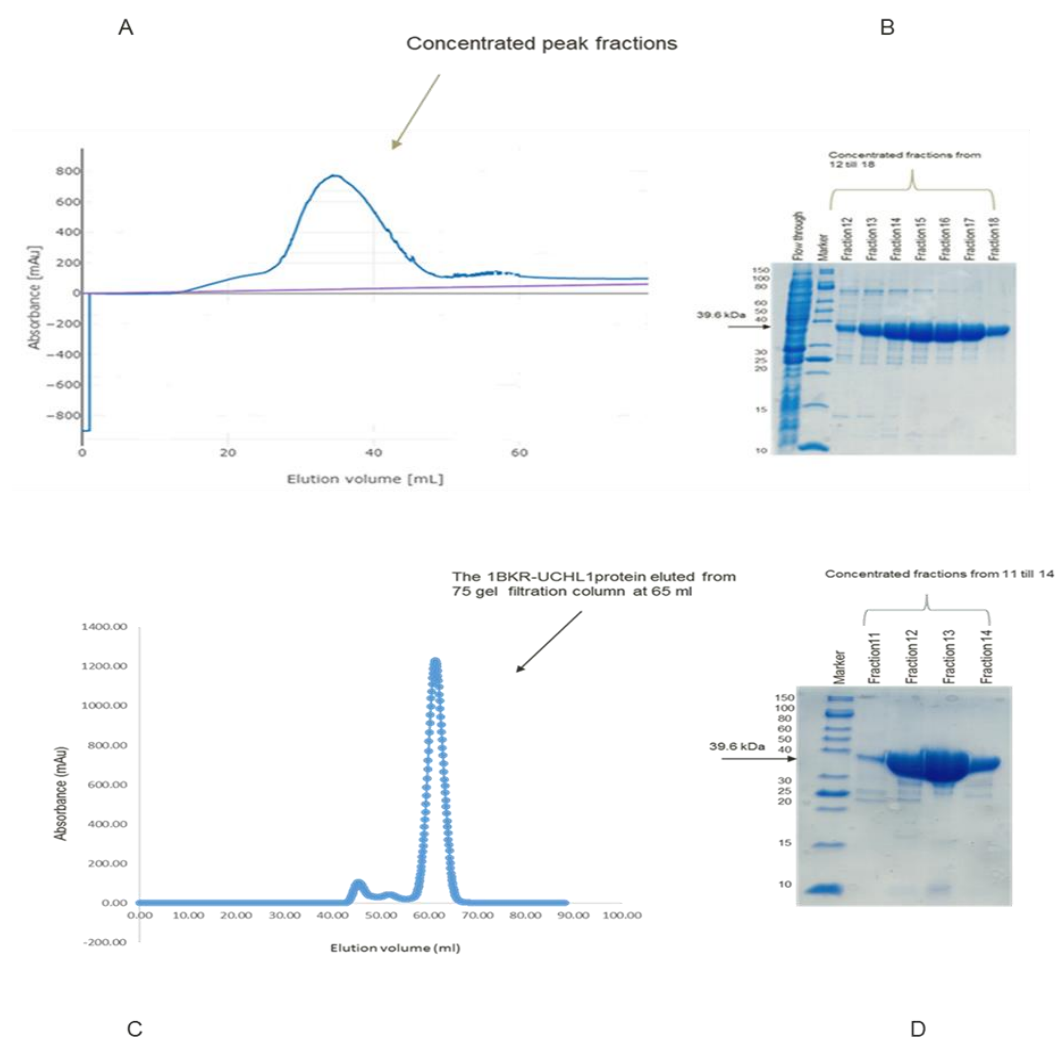


Figure 31. A: Affinity chromatography of 1BKR-UCHL1. B: SDS-PAGE gel from Ni column showed 1BKR UCHL1 has 39.6 kDa. C: Chromatogram of 1BKR-UCHL1 after gel Filtration purification. D: SDS-PAGE gels showed a band at around 40 kDa, which is analogous to that of 1BKR UCHL1.

4.1.3.1.2 2GKG-UCHL1 protein purification

Nickel column purification of 2GKG-UCHL1 protein revealed a broad single peak (Figure 32 A) indicating a high protein yield which similar to 1BKR-UCHL1 protein results. Analysis of the fractions-under the peak on SDS-PAGE revealed that the protein had bound tightly to the nickel column and the protein was soluble. The SDS-PAGE gel also showed overloading of the protein samples (Figure 32 B), as there was a band between the 40 and 25 kDa marker bands, which is analogous to the predicted size of the expressed protein. Analysis also revealed some contaminating proteins which required another purification step before crystallization trials, as protein should be 90 percent pure before crystallization trials (231).

The gel filtration as second step of purification revealed two peaks (Figure 32 C). The first peak was the void volume and the second peak contained the 2GKG-UCHL1 protein as indicated by standard curve prediction of molecular weight of about 40.4 kDa. This protein eluted at around 64 mL indicating that there was no aggregation of 2GKG-UCHL1 protein, as the protein eluted in the correct elution volume not in the void volume (219). SDS-PAGE gel analysis confirmed the protein purity by absence of bands except the specific target protein band in the correct size (Figure 32 D). The final yield after concentration of the purified protein was 25 mg/L being higher than that observed for 1BKR-UCHL protein. This result suggests that the presence of 2GKG had a more pronounced effect on producing the UCHL1 soluble protein than the 1BKR tag.

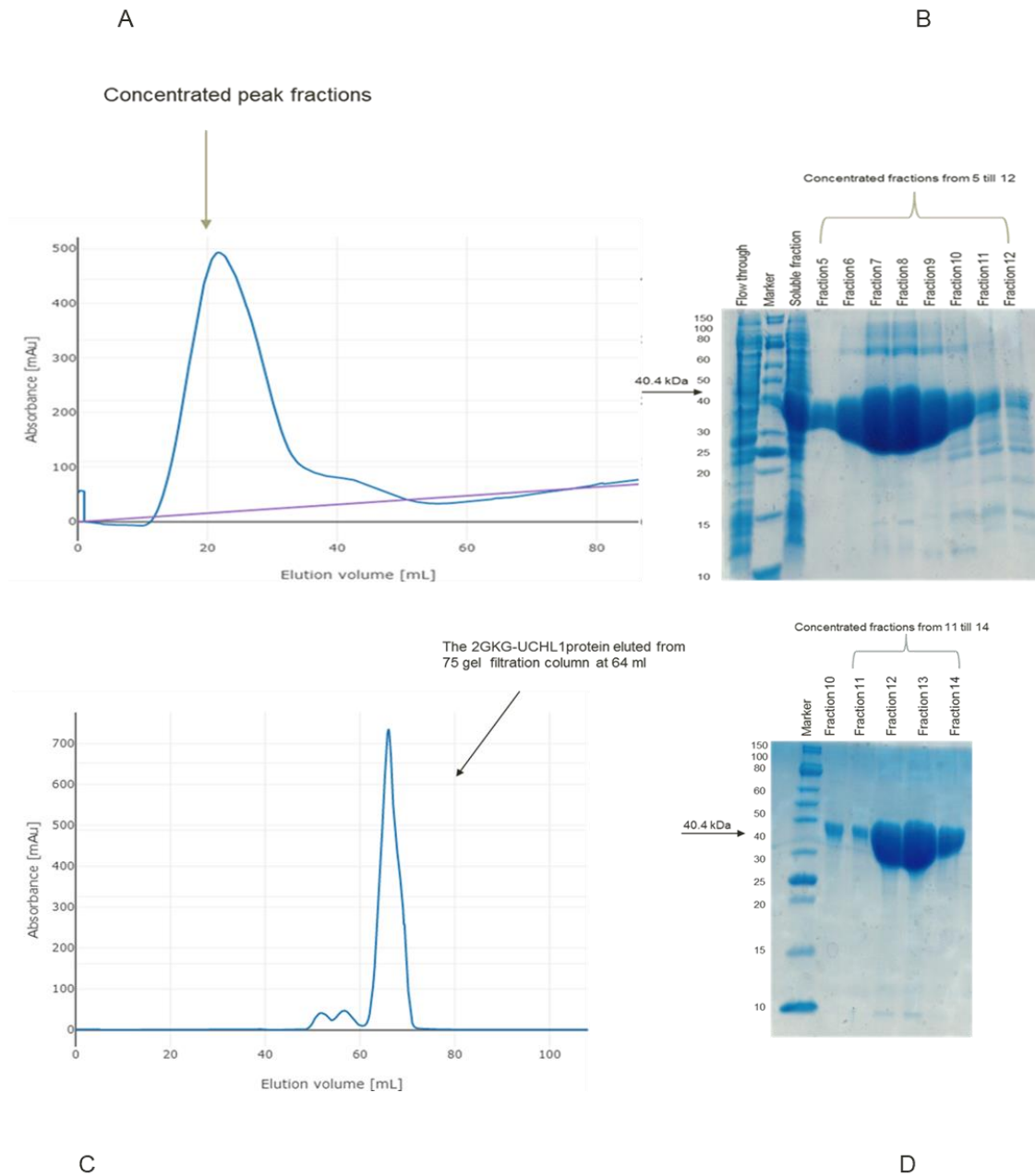


Figure 32. A: Chromatogram of the affinity chromatography showed a broad peak which contains 2GKG-UCHL1 protein. **B:** SDS-PAGE gel from Ni column showed 2GKG-UCHL1. **C:** Gel filtration chromatogram of 2GKG-UCHL1. **D:** Analysis of the peak fractions on SDS-PAGE gels revealed a band of around 40 kDa, this band was the correct predicted size for 2GKG-UCHL1.

4.1.3.1.3 MBP-UCHL1 protein purification

Purification using Nickel affinity chromatography of MBP-UCHL1 revealed one peak present on the chromatogram (Figure 33 A). Analysis of the peak fractions on a SDS-PAGE gel revealed the presence of the MBP-UCHL1 expressed protein as a distinct single band of around 60-66 kDa (Figure 33 B).

Gel filtration purification was done as a final purification step showing a small void volume peak and a single peak. The single peak represents the target protein being this protein eluted at 88 mL (Figure 33 C). This volume corresponded to the elution profile of a 66.2 kDa protein by using the standard curve of the gel filtration column Superdex 200 (16/60) (232). SDS PAGE gel analysis of the peak fractions indicated that the MBP-UCHL1 protein was present as a purified protein by distinct single band at about 66.2 kDa (Figure 33 D). After collection and concentration, the total yield of MBP-UCHL1 protein was 27 mg/L which was more than that observed for 2GKG and 1BKR tags. This data suggests that the MBP tag had the highest effect on the UCHL1 production.

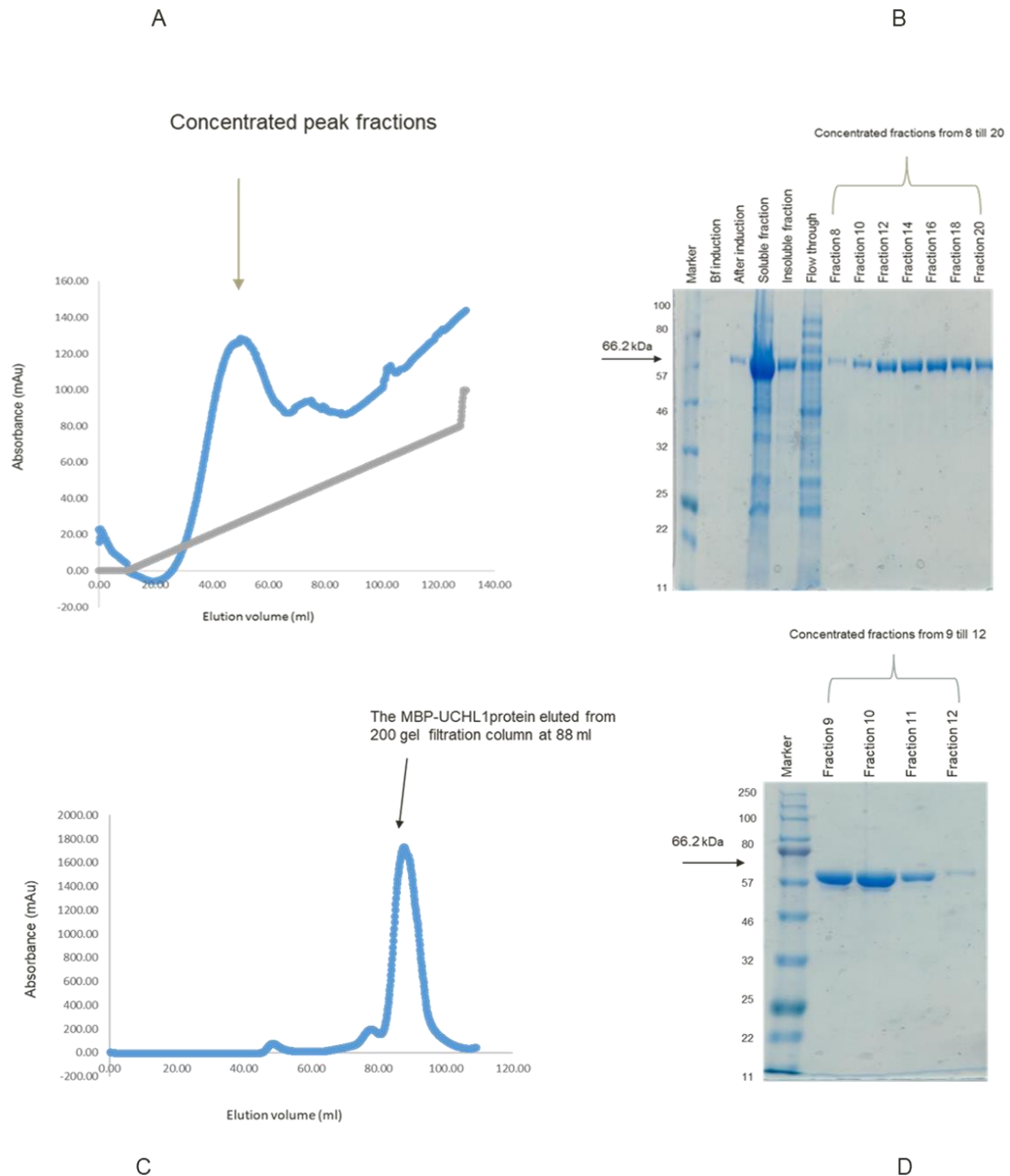


Figure 33. A: The chromatogram of affinity chromatography of MBP-UCHL1. Chromatogram showed a prominent peak which contained the target protein. B: Analysis of the peak fractions and insoluble pellets on SDS-PAGE gel revealed protein bands which corresponded to the size of MBP-UCHL1 of approximately 66.2 kDa. C: The chromatogram of gel filtration purification of MBP-UCHL1. D: SDS-PAGE gel result showed bands that corresponded to MBP-UCHL1 (66.2 kDa).

4.1.4 Investigation of UCHL1 protein- solubility tag fusions using small Angle X ray scattering (SAXS)

SAXS analysis gives important information with regards to how fusion tags affect the UCHL1 protein in terms of protein aggregation, folding, size and multiple conformation flexibility (233). SAXS analysis of 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1 proteins revealed that these proteins were monodisperse without any indication of particle aggregation, since the proteins scattered were in a straight line indicating the normal distribution for folded non aggregated proteins according to a Guinier plot (Figure 34 A,B,C) (234). The Guinier plot estimates the radius of gyration (R_g) of 1BKR-UCHL1 (24.6 Å), 2GKG-UCHL1 (24.8 Å) and MBP-UCHL1 (30.8 Å).

Plotting dimensionless Kratky plots also confirmed that all the proteins were folded with no aggregation, as all the proteins were characterised by dimensionless Kratky plots bell-shaped curve. This is a characteristic of a folded protein however, some flexibility was seen as the curves did not fully converge to the x axis (235). The 1BKR-UCHL1 and 2GKG-UCHL1 dimensionless Kratky plots showed slightly sharper but also broadened peaks than that observed for the MBP-UCHL1 which was characterised by having more of a profile of a multiple domain protein. These curves were characteristic of the proteins being folded with some flexibility domains (236). Pair-Distribution Function $P(r)$ plots of 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1 determined that the proteins were folded and had a smooth distribution function (Figures 36 A, B, C). The $P(r)$ plots of 1BKR-UCHL1 and 2GKG-UCHL1 had shoulders in their bell-shape consistent with the two domain nature of the fusion proteins (237). The maximum linear dimension (D_{max}) of 1BKR-UCHL1 was 77 Å

which correlated well with the predicted D max from a model using Pymol software which was 75.7 Å (Figure 37 A). The maximum linear dimension (Dmax) of 2GKG-UCHL1 was 78 Å which was similar to the estimates of D max using Pymol software which was 79.2 Å (Figure 37 B). The maximum linear dimension (Dmax) of MBP-UCHL1 was 89.0 Å, which was analogous to the predicted of D max of a model using Pymol software which was 86 Å (Figure 37 C).

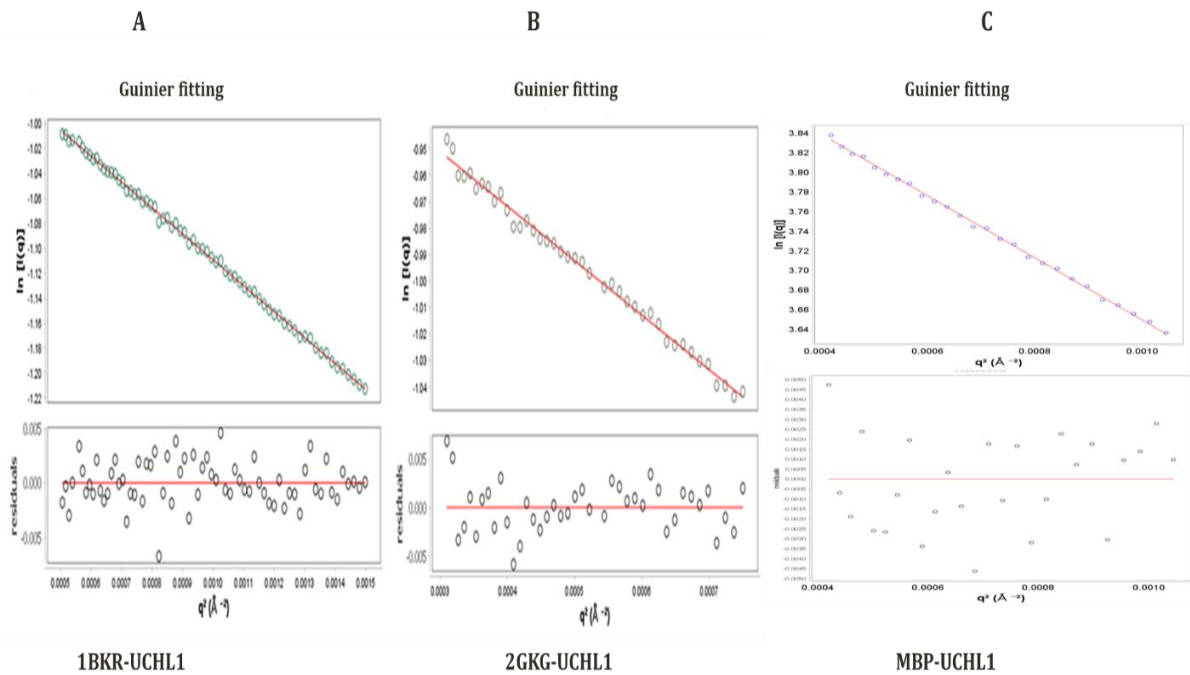


Figure 34.:A: The Guinier plot of the 1BKR UCHL1 protein. B: Guinier plot of the 2GKG UCHL1 protein. C: Guinier plot of MBP-UCHL1 protein.

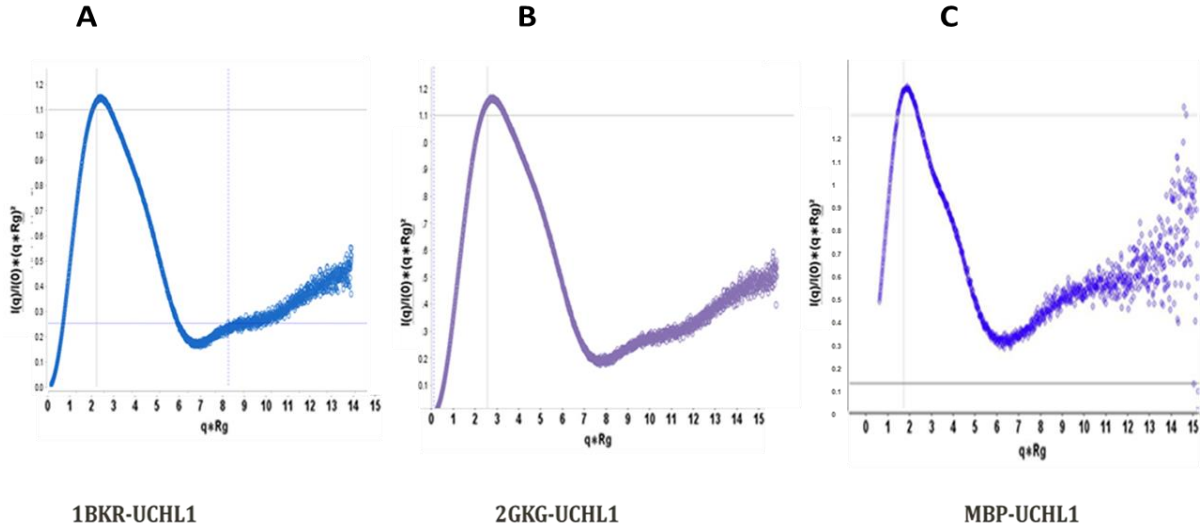


Figure 35. A: Dimensionless Kratky plot of 1BKR-UCHL1 protein, B: Dimensionless Kratky plot of 2GKG-UCHL1 protein. C: Dimensionless Kratky plot of MBP-UCHL1 protein.

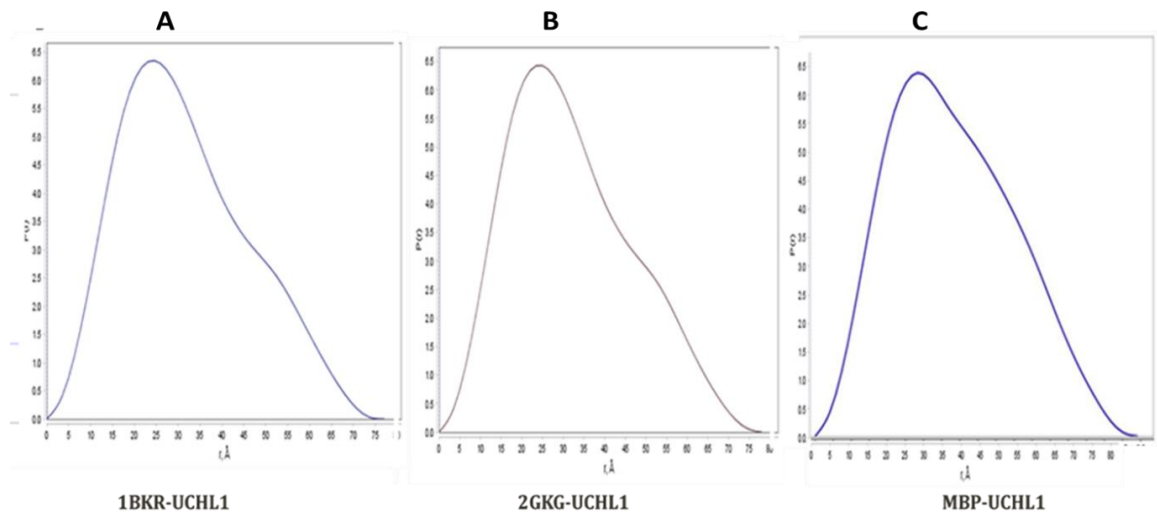


Figure 36. A: Pair-Distribution Function $P(r)$ plot of 1BKR-UCHL1 protein, B: Pair-Distribution Function $P(r)$ plot of 2GKG-UCHL1 protein. C: Pair-Distribution Function $P(r)$ plot of MBP-UCHL1 protein.

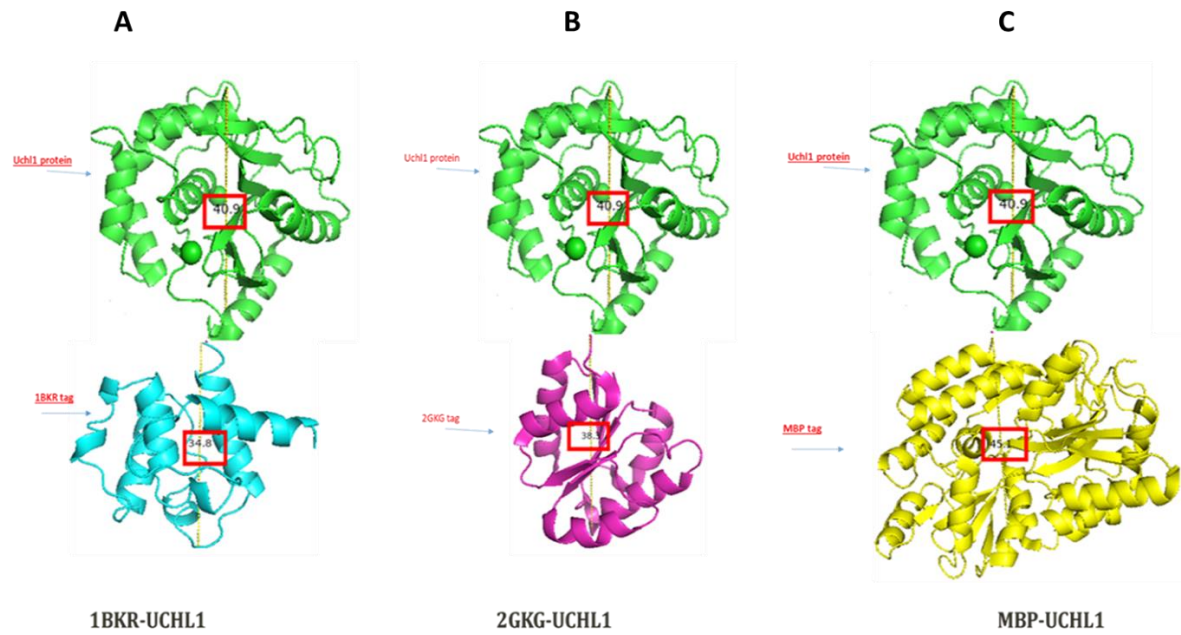


Figure 37. A: The prediction of the maximum linear dimension (Dmax) by pymol software of models of the 1BKR-Uchl1 proteins from their crystal structures which is equal 75.7 Å. **B:** The prediction of the maximum linear dimension (Dmax) by pymol software of the 2GKG-Uchl1 protein from their crystal structures which was equal to 79.2 Å. **C:** The prediction of the maximum linear dimension (Dmax) by pymol software of the MBP-Uchl1 protein from their crystal structures which was equal to 86 Å.

4.1.5 Crystallisation trials of UCHL1 protein- solubility tag fusions

Crystallisation trials of 1BKR-UCHL1, 2GKG-UCHL1 and MBP-UCHL1 proteins were performed using Morpheus screen, JCSG+, PACT premier screen, PEG Screen (Grid screen) and Ammonium sulphate (Grid screen). The concentrations used were 15 and 30 mg/mL protein. However, no crystals were formed after one week. The screens with different concentrations after one week seemed similar to each other with only data from the 30 mg/mL after one week experiments presented in the tables 8, 9 and 10.

Crystallization trials: 1BKR-UCHL1 protein 30 mg/mL			
screens	Clear	Light precipitate	Heavy precipitate
Amm sulphate	29	57	10
JCSG plus	11	62	23
PACT Premier	10	65	21
Morpheus	18	21	61
PEG Screen	19	44	33

Table 8. Results of the crystallisation trials using different crystallisation screens for 30 mg/mL 1BKR-UCHL1 protein after one week

Crystallization trials: 2GKG-UCHL1 protein 30 mg/mL			
screens	Clear	Light precipitate	Heavy precipitate
AMM. sulphate	20	51	25
JCSG plus	7	51	38
PACT	17	48	31
Morpheus	8	28	60
PEG Screen	27	23	46

Table 9. Results of the crystallisation trials using different crystallisation screens for 30 mg/mL 2GKG-UCHL1 protein after one week

Crystallization trials: MBP-UCHL1 protein 30 mg/mL			
screens	Clear	Light precipitate	Heavy precipitate
AMM. Sulphate	20	41	35
JCSG plus	32	57	7
PACT	17	38	41
Morpheus	20	19	57
PEG Screen	25	29	42

Table 10. Results of the crystallisation trials using different crystallisation screens for 30 mg/mL MBP-UCHL1 protein after one week

4.2 Results of the investigation of UCHL3-tag fusions

This section investigated another model protein with known structure when fused with the two novel tags under investigation and the MBP tag. The second model protein chosen was the UCHL3 protein. This section compares the results of the UCHL3 enzyme fused with each tag in small and large-scale cultures.

4.2.1 Cloning of 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3 constructs

Three constructs were generated 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3. The UCHL3 gene was successfully cloned into the 1BKR-pRSF-13, 2GKG-pRSF-13 and MBP- pRSF-13 vectors as described for the previous model protein. The sequencing results confirmed that the 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3 constructs had been successfully engineered.

4.2.2 Expression, solubility and small-scale purification of 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3

These experiments were done to evaluate if the tags can express the UCHL3 protein and to make a qualitative comparison between 1BKR, 2GkG and MBP tags fused with UCHL3 protein by auto induction protocol at a small scale (100 mL).

Analysis of non-induced and induced fractions from bacterial cultures of the relevant proteins on SDS-PAGE revealed that all the proteins had been expressed indicated by the existence of an extra protein band in the induced fractions (Figure 38 A). All induced fractions bands, for each protein, were of the correct size. The predicted

molecular weight of MBP-UCHL3 was 67 kDa, 2GKG-UCHL3 was 40.4 kDa and 1BKR-UCHL3 was 39.4 kDa. The SDS-PAGE result showed that the bands of the soluble fractions of 2GKG-UCHL3 and 1BKR-UCHL3 were more intense than MBP-UCHL3 (Figure 38 B). Analysis of the samples after purification on a SDS-PAGE gel revealed that the three proteins 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3 had been successfully purified. SDS-PAGE analysis indicated that the protein band of the 2GKG tag was more intense than that observed for the other tags (Figure 38 C).

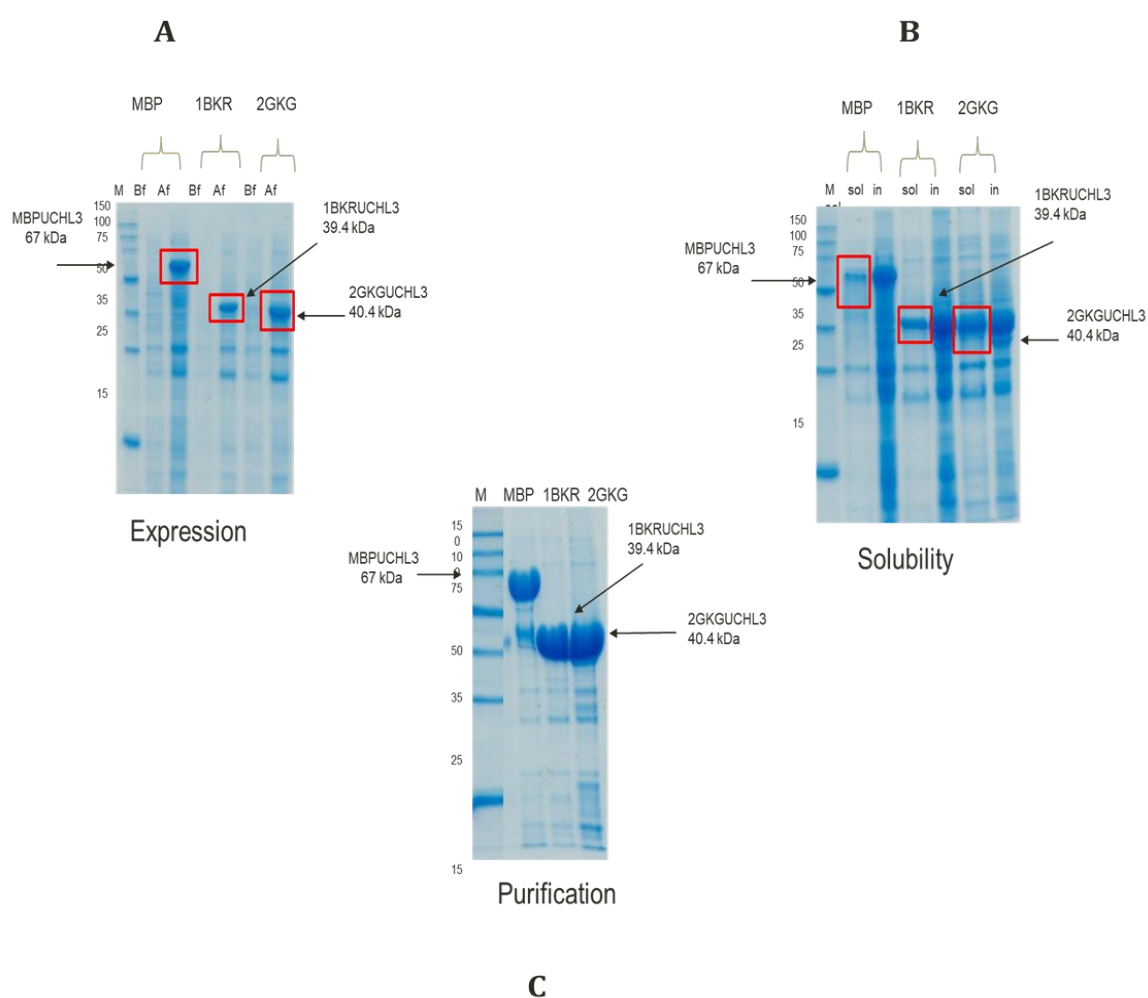


Figure 38: A: SDS-PAGE gel of expression test of the three proteins 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3. B: SDS-PAGE gel of solubility test of the three proteins 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3. C: SDS-PAGE gel of purification test of the three proteins 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3.

4.2.3 Large-scale expression and purification of UCHL3 protein-solubility tag fusions

The aim of this experiment was a quantitative comparison between 1BKR-UCHL3, 2GKG- UCHL3 and MBP-UCHL3 proteins at a larger volume scale and to produce proteins at a high yield for subsequent crystallization trials. Three solubility-tagged constructs of the UCHL3 protein were expressed in two litres cultures by IPTG protocol. Using nickel affinity and then gel filtration chromatography, the three proteins were purified.

4.2.3.1 1BKR-UCHL3 purification

Nickel affinity chromatography was employed to purify 1BKR-UCHL3. The purification analysis revealed a large single peak indicating a high amount of protein was purified (Figure 39 A). The SDS-PAGE gel showed the protein was well expressed and tightly bound to the nickel column, as no protein was observed in fraction from the flow through (Figure 39 B). The SDS-PAGE result demonstrated that there was an intense band between the 35 and 45 kDa marker bands, which is analogous to the predicted size of the 1BKR-UCHL3 protein. Size exclusion purification was performed for an assessment of protein stability (238). The size exclusion purification of 1BKR-UCHL3 was characterised by the presence of two peaks. The void volume was the first peak (Figure 39 C). The target protein was in the second peak as eluted at the expected elution volume of 65 mL from the gel filtration (Superdex 75 (16/60) column. This volume correlates with the standard curve prediction of molecular weight of this protein of about 39.4 kDa. The results showed no aggregation of the 1BKR-UCHL3 protein.

The SDS-PAGE gel results on the protein purity indicated that there had been little degradation (Figure 39 D) and total protein yield was 6.91 mg/L culture.

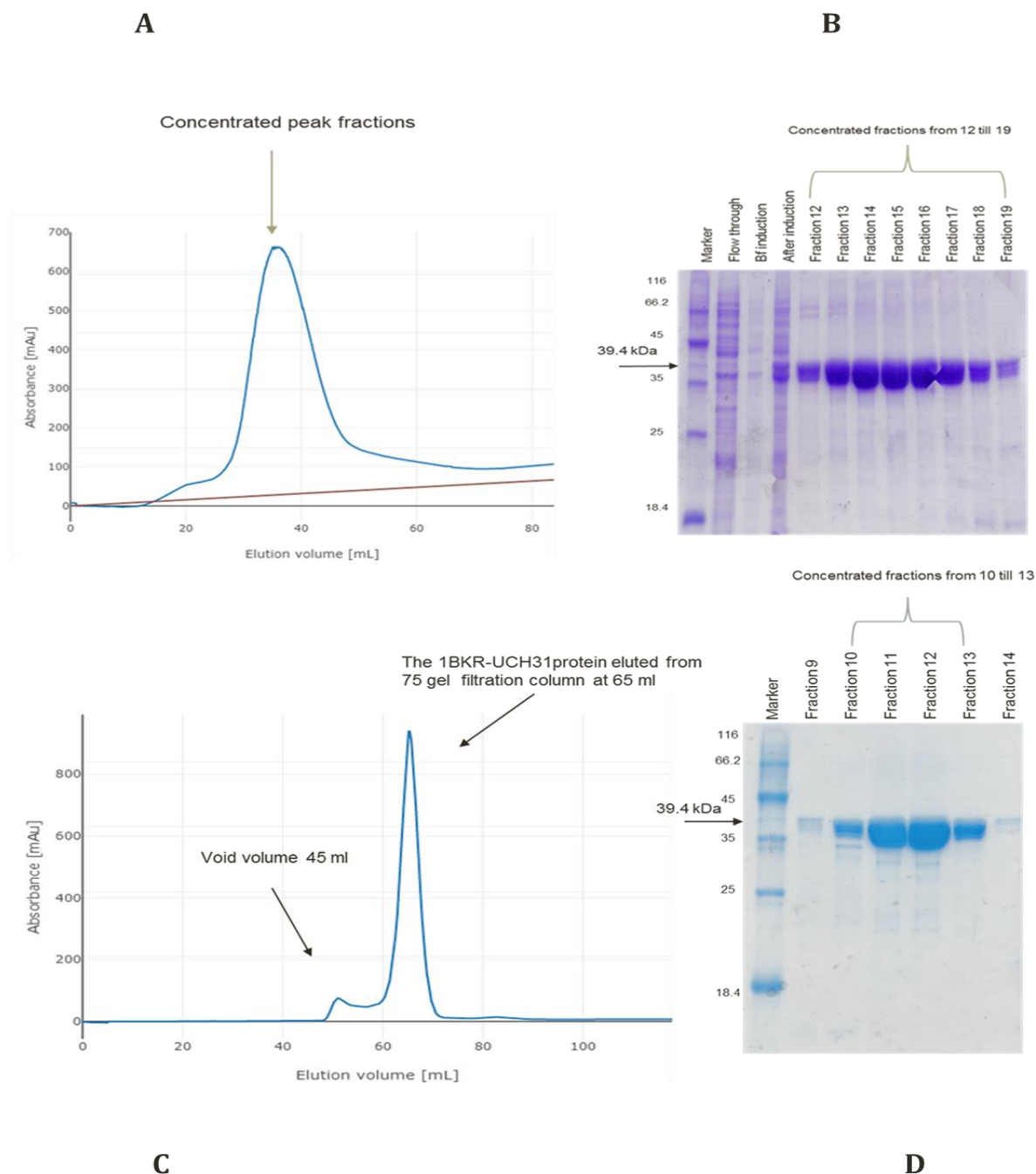


Figure 39.A:Affinity chromatography and SDS-PAGE gel of 1BKR-UCHL3.B: SDS-PAGE gel from Ni column. C: Chromatogram of 1BKR-UCHL3 gel Filtration purification. D: SDS-PAGE gels represented a band at around 40 kDa which is a 1BKR UCHL3.

4.2.3.2 2GKG-UCHL3 purification

2GKG-UCHL3 was purified using the same methods as employed for 1BKR-UCHL3. The purification using a Nickel column revealed a prominent broad single peak which indicates protein with a high yield was obtained (Figure 40 A). The peak fractions analysis by SDS PAGE indicated that the 2GKG-UCHL3 protein had entirely bound to the nickel column, as there is no band in the flow through fraction of the correct protein size. The SDS PAGE gel showed also an intense protein band between protein standard 35 kDa and 45 kDa marker bands (Figure 40 B).

Gel filtration purification chromatogram was characterised by the presence of two peaks. The first peak was the void volume and the second peak contained the target protein (Figure 40 C). A standard curve suggested of molecular weight of about 40.4 kDa, which corresponds to an elution volume of 64 mL (239). This result showed some aggregation of the 2GKG-UCHL3 protein as the void volume peak was not totally separated from the peak of the protein target and this affected the final yield. The purified fractions analysis by SDS-PAGE gel revealed a distinct protein band of the expected size with little contaminating protein bands and the protein was considered pure enough for crystallization trials (Figure 40 D). The total protein yield was 7.6 mg/L culture, which is more than that observed for the 1BKR tag fusion suggesting the presence of the 2GKG tag had a better effect on UCHL3 protein production yield.

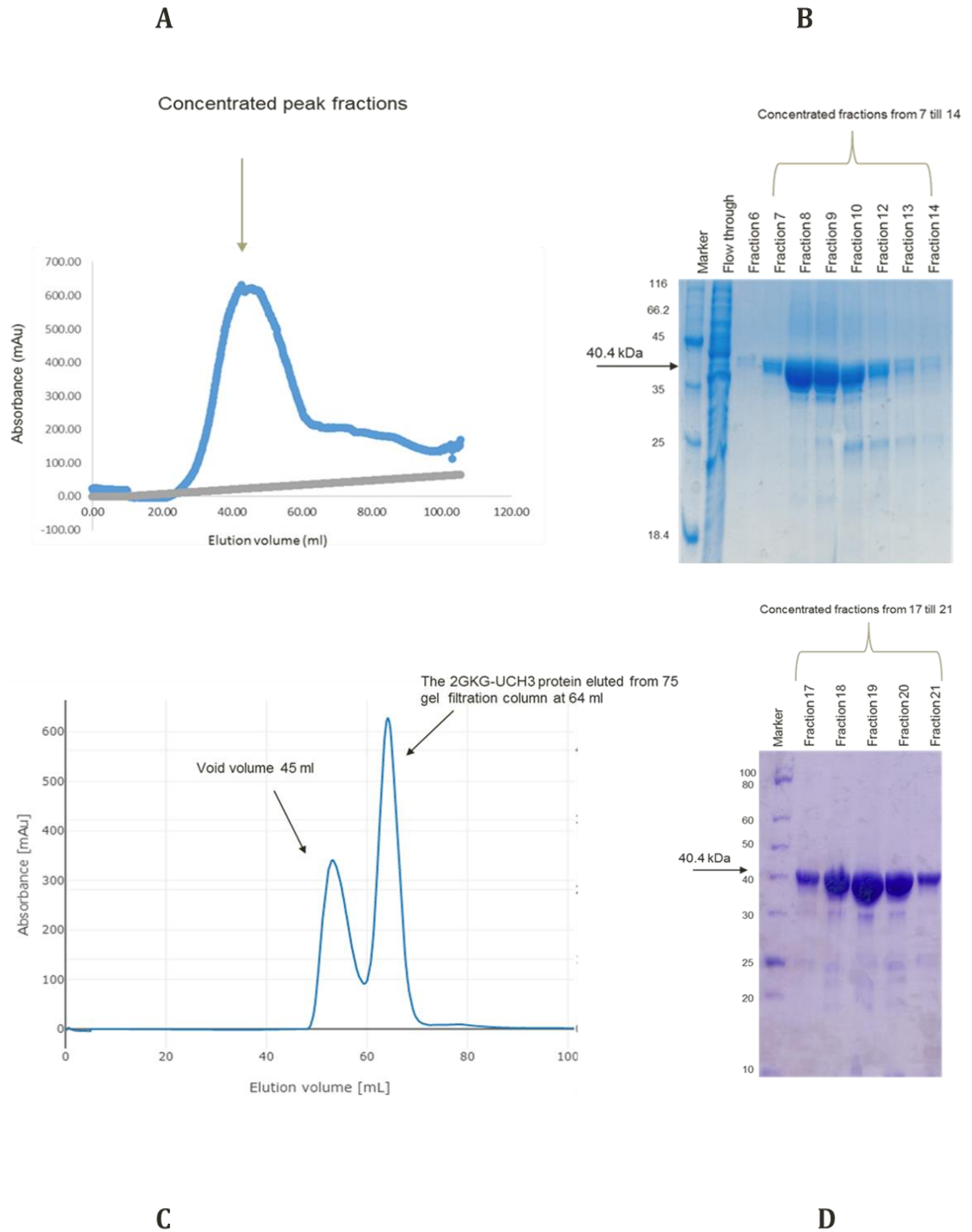


Figure 40. A. The affinity column chromatogram. B: SDS-PAGE gel from a Ni column run revealed a band at 40 kDa analogous to the predicted size of 2GKG-UCHL3 with some contaminating proteins. C: Chromatogram of 2GKG-UCHL3 gel filtration purification. D: SDS-PAGE gels revealed a band at around 40 kDa that is 2GKG-UCHL3.

4.2.3.3 MBP-UCHL3 purification

The MBP-UCHL3 protein was purified using the same methods as described for the 1BKR-UCHL3 and 2GKG-UCHL3 fusions. The chromatogram from the Nickel affinity column purification showed a single broad peak which contained the target protein (Figure 41 A). Analysis of purified peak fractions by SDS-PAGE gel revealed that the protein had been expressed and was present as a clear single band of around 67 kDa (Figure 41 B).

The chromatogram of gel filtration purification showed two peaks (Figure 41 C). The first peak could be the protein being present as a dimer but not the void oligomer, as the standard curve gel filtration column Superdex 200 (16/60) represented. The high peak contained the protein target, which eluted at 87.8 mL which corresponds to the elution of a 67 kDa protein (240). The peak fraction analysis by SDS PAGE gel revealed that a pure protein had been obtained (Figure 41 D). SDS PAGE gel revealed a clear single band at approximately 67 kDa. The peak fractions were combined and concentrated for crystallization studies. After collection and concentration, the total protein yield was 7.5 mg/L culture, which is very close to the 2GKG-UCHL1 soluble yield, but higher than 1BKR-UCHL1, which suggests the presence of MBP and 2GKG had a bigger effect than the presence of the 1BKR tag with the UCHL3 protein.

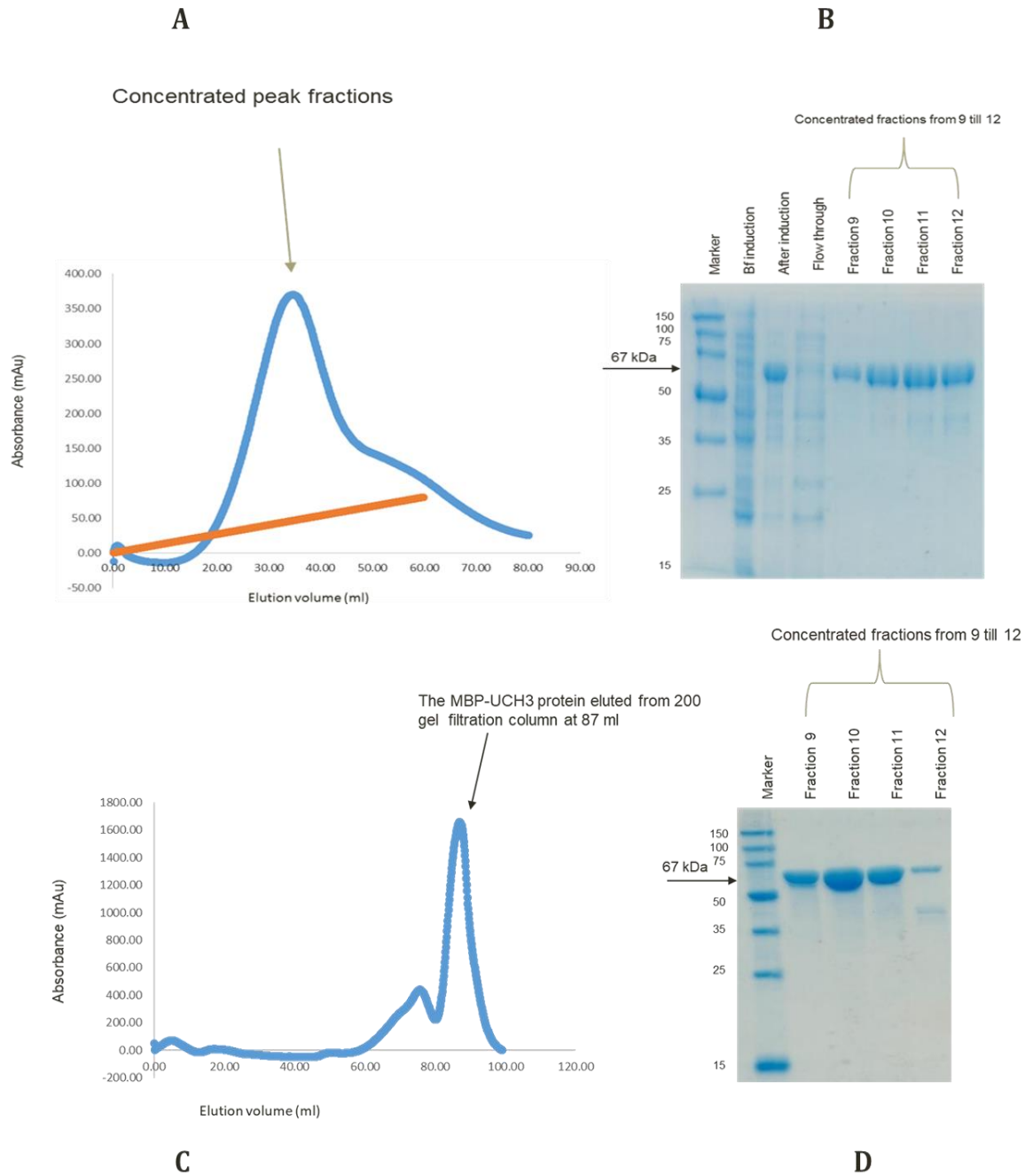


Figure 41. A: The affinity chromatography chromatogram of MBP-UCHL3. B: SDS-PAGE gel from Ni column showed MBP-UCHL1 at approximately 67 kDa C: Chromatogram of MBP-UCHL3 gel filtration purification. D: SDS-PAGE gels represented a band at about 67 kDa, this band is the MBP-UCHL3 protein.

4.2.4 Analysis of UCHL3 protein- solubility tag fusions using small Angle X ray scattering (SAXS)

These experiments were done as described for the UCHL1 protein to evaluate the effect of the novel tags on the UCHL3 protein. These assays determined whether the proteins aggregated or folded correctly and provided details on the size and multiple conformations of a protein flexibility, which gave valuable information to know how the presence of fusion tags affected the fusion protein (241). The results of 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3 proteins revealed that they were found as a monodisperse protein, as the protein scattered in a straight line with a normal distribution, as represented in a Guinier plot (Figure 42 A, B, and C). The Guinier plot determined the radius of gyration (R_g) of these proteins 1BKR-UCHL3 (25.7 Å), 2GKG-UCHL3 (25.91 Å) and MBP-UCHL3 (29.8Å). The dimensionless Kratky plots of the 1BKR-UCHL3 protein showed the peak was a regular bell-shaped curve, which indicated the protein was relatively globular, folded and not aggregated (Figure 43 A, B, and C). The peak of the 2GKG-UCHL3 showed that this protein was less flexible than other proteins as the curve converged to the x axis more than that observed with the other two proteins (Figure 43 B) (242). Pair-Distribution Function $P(r)$ plots of the three proteins confirmed the same results of dimensionless Kratky plots, and revealed that the proteins were folded and had a smooth distribution function. A relatively regular bell-shaped curve was observed for 1BKR-UCHL3 whereas the two other fusions displayed a more two-domain characteristic $P(r)$ function (243) (Figures 44 A,B,C). The maximum linear dimension (D_{max}) of 1BKR-UCHL3 was 108.5 Å while the prediction of D_{max} by pymol software was only 70.9 Å (Figure 45 A). The maximum linear dimension (D_{max}) of 2GKG -UCHL3 was 85Å while the prediction of D_{max} of a

model in pymol was 77Å (Figure 45 B). The maximum linear dimension (Dmax) of MBP-UCHL3 was 91.0 Å, with a prediction D max using pymol software was 83.6Å (Figure 45C)

These results suggest that the proteins were present as a dimer in the solution or adopt a more elongated conformation than modelled as the Dmax values obtained by pymol software is less than that obtained using Pair-Distribution Function P(r) prediction (244).

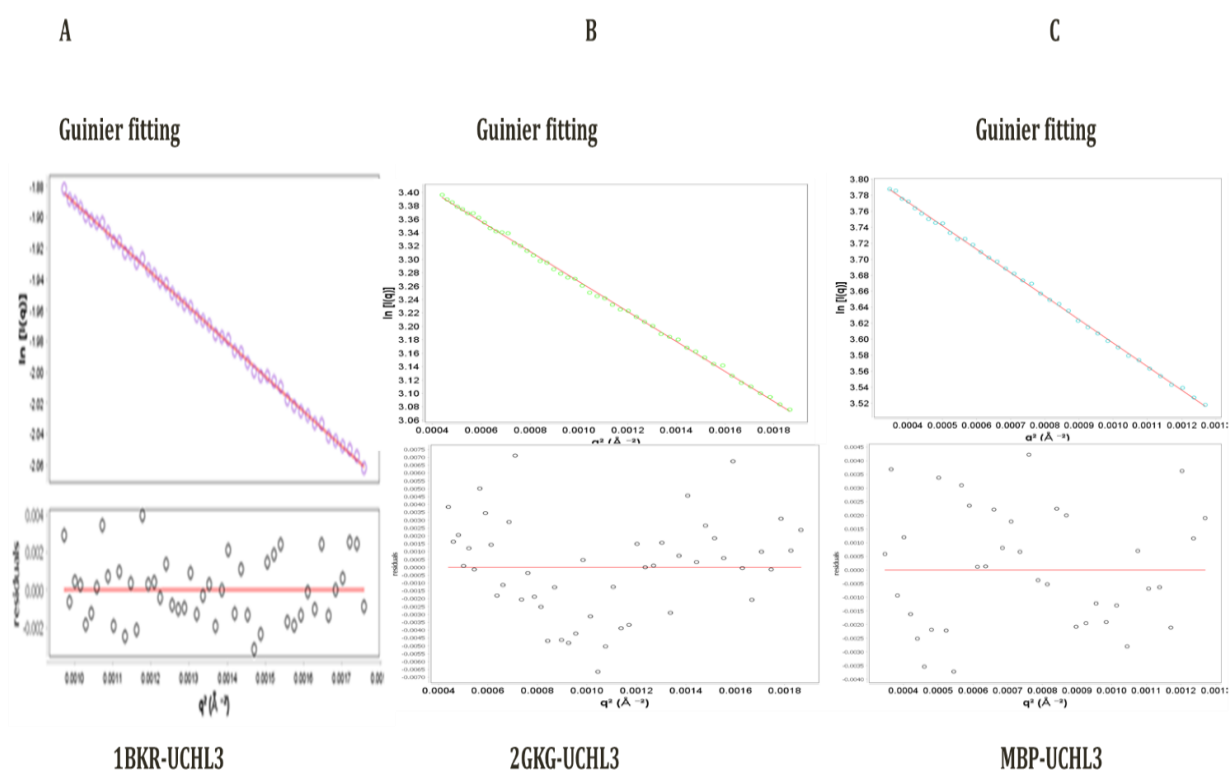


Figure 42:A: The Guinier plot of 1BKR-UCHL3 protein. B: The Guinier plot of 2GKG-UCHL3 protein. C: The Guinier plot of MBP-UCHL3 protein.

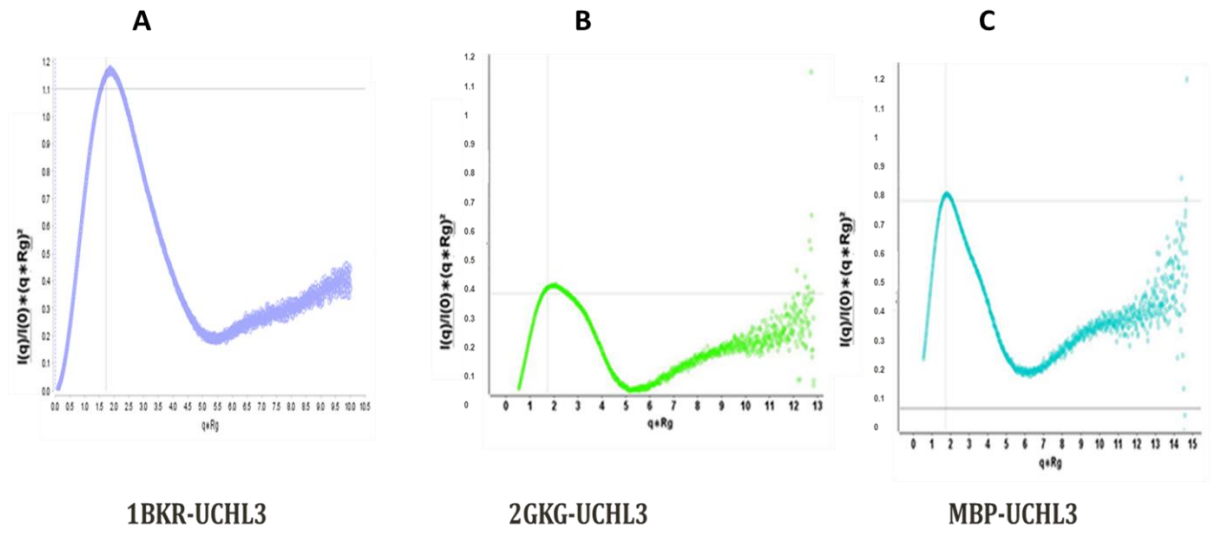


Figure 43. A: Dimensionless Kratky plot of 1BKR-UCL3 protein. B: Dimensionless Kratky plot 2GKG-UCL3 protein. C: Dimensionless Kratky plot of MBP-UCL3 protein.

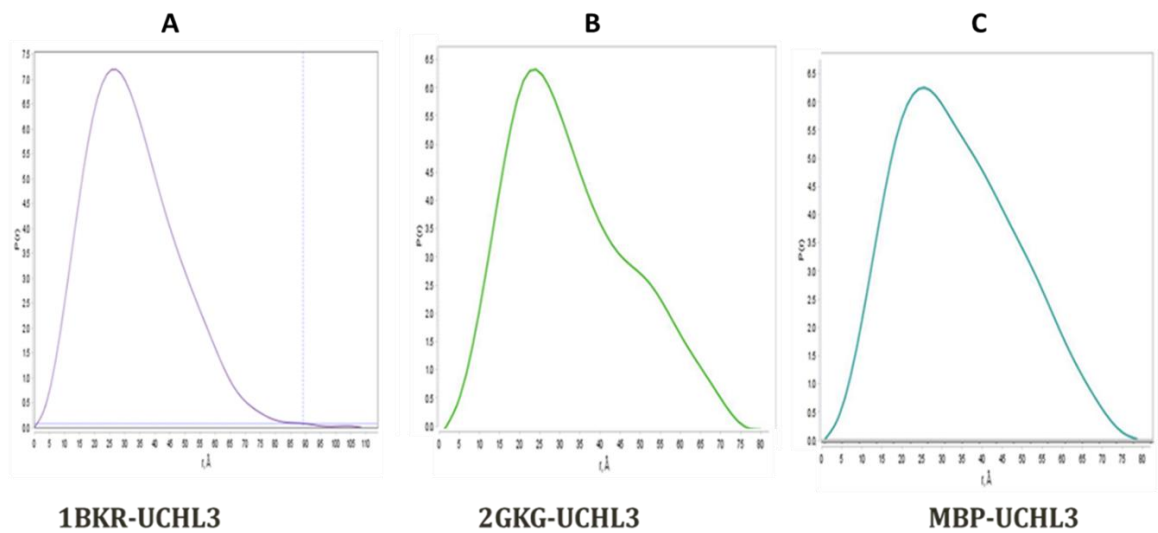


Figure 44. A: Pair-Distribution Function $P(r)$ plot of 1BKR-UCL3 protein. B: Pair-Distribution Function $P(r)$ plot of 2GKG-UCL3 protein. C: Pair-Distribution Function $P(r)$ plot of MBP-UCL3.

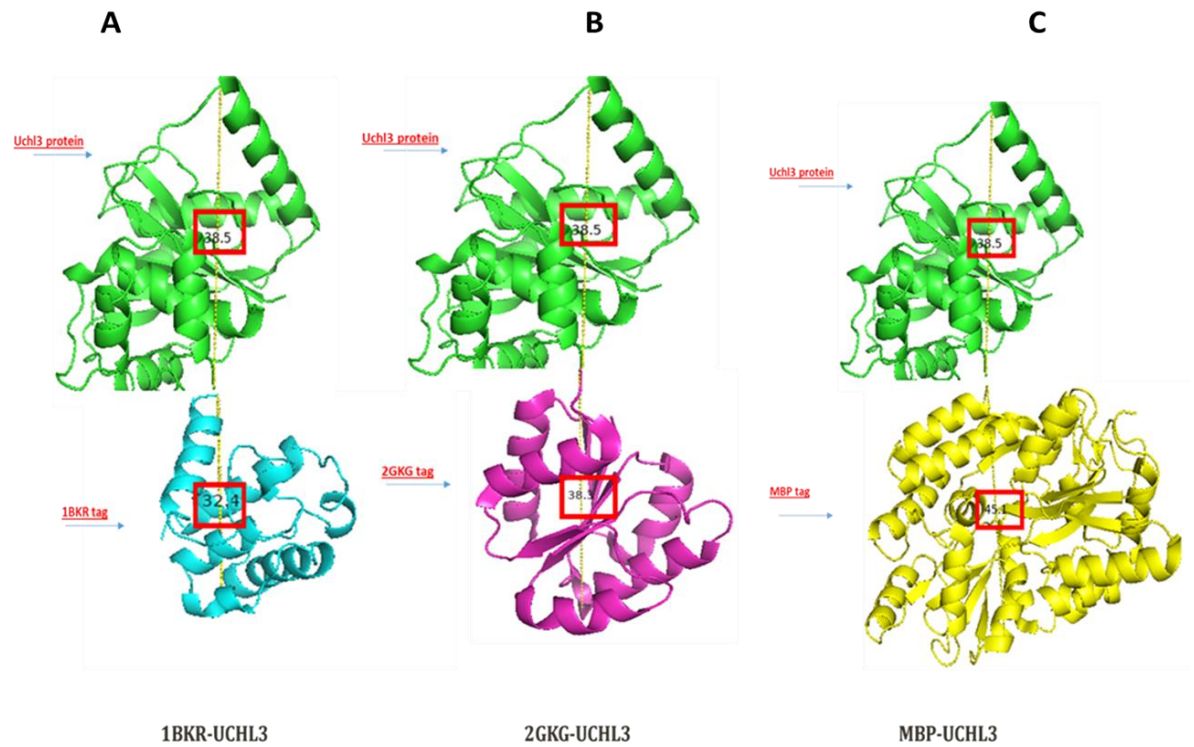


Figure 45. The prediction of the maximum linear dimension (Dmax) by Pymol software of the 1BKR-Uchl3 protein from their crystal structures, which is equal to 70.9 Å. B: The prediction of the maximum linear dimension (Dmax) by pymol software of the 2GKG and Uchl3 proteins from their crystal structures which was equal to 77 Å. C: The prediction of the maximum linear dimension (Dmax) by pymol software of the MBP and Uchl3 proteins from their crystal structures which is equal 83.6 Å.

4.2.5 Crystallisation trials of UCHL3 protein-solubility tag fusions

The crystallisation trials of proteins 1BKR-UCHL3, 2GKG-UCHL3 and MBP-UCHL3 were performed using 10, 15 and 20 mg/mL protein concentrations using the following screens: Morpheus, Ammonium sulphate, PACT Premier, PEG screen, and JCSG plus. Each protein was incubated for 1 week at 20°C and no crystals were obtained from 1BKR-UCHL3 and MBP-UCHL3 proteins (Table 12). On the other hand, a small crystal was obtained of the 2GKG-UCHL3 protein in the PEG screen (Figure 39) but no diffraction was obtained. The crystallisation screen result after one week from the three proteins at 15 mg/mL are shown in table 11, 12 and 13. Results from crystallization trials using different protein concentrations (10 and 20 mg/mL) gave the similar results as presented in three tables 11, 12 and 13.

Crystallization trials: 1BKR-UCHL3 protein 15 mg/mL			
screens	Clear	Light precipitate	Heavy precipitate
AMM. sulphate	36	53	7
JCSG plus	29	37	30
PACT	23	49	24
Morpheus	15	35	46
PEG Screen	29	34	33

Table 11. Results of crystallisations using different crystal screens for 15 mg/mL 1BKR-UCHL3 protein after one week.



Figure 46. The crystal from 2GKG -UCHL3 but no diffraction was obtained. The condition was 0.2 M Calcium chloride dihydrate 0.1 M Tris 20 % w/v PEG 6000, pH=8.

Crystallization trials: 2GKG-UCHL3 protein 15 mg/mL			
screens	Clear	Light precipitate	Heavy precipitate
Amm sulphate	48	43	5
JCSG plus	38	41	17
PACT	19	54	23
Morpheus	17	27	52
PEG Screen	38	45	13

Table 12. Results of crystallisations using different crystal screens for 15 mg/mL 2GKG-UCHL1 protein after one week.

Crystallization trials: MBP-UCHL3 protein 15 mg/ml			
screens	Clear	Light precipitate	Heavy precipitate
AMM.sulphate	26	34	36
JCSG plus	23	35	38
PACT	27	43	26
Morpheus	21	30	45
PEG Screen	41	17	38

Table 13. Results of the crystallisation trials using different crystal screens for 15 mg/mL MBP-UCHL1 protein after one week.

4.3 Discussion of the results of UCHL1 and UCHL3 proteins with the tags

This chapter results showed the expression, solubility and purification of UCHL1 and UCHL3 proteins when fused to each of 2GKG, 1BKR and MBP tags. UCHL1 and UCHL3 proteins were selected in order to potentially crystallize these proteins in different space groups (crystal packing) that may be better suited for compound soaking. In addition, the proteins were relatively globular and their structures were known, which made it more straight forward to design the linkers.

The first experiment was done in small scale utilizing qualitative SDS-PAGE analysis as a reliable indication of the predicted quantities of protein production in large expressions (245). The results of small-scale comparisons showed that the expression, solubility and purification of MBP and 2GKG tags were not different from each other and were better than that observed with a1BKR tag. These results indicated that the 2GKG tag gave almost similar results to that observed with the well-described best fusion partner MBP in terms of soluble protein expression.

The results of the large-scale experiment were consistent with the small-scale experiment results but in terms of the amount of protein yield, as the data showed that the MBP and 2GKG tags yielded nearly the same yield and purified proteins. Both were better than the yield of a 1BKR tag. (Table 14,15) (Figure 48,49).

The expression and purification results of small and large scale of both 1BKR and 2GKG had successfully promoted the correct folding and solubility of UCHL1 and UCHL3 proteins this supports our hypothesis that 1BKR and 2GKG could be used as

solubility tags. Additionally, the small and large scale results of our novel tags were the nearly similar results that observed with Fh8 tag (8 kDa) when comparing with MBP tag with different chosen model proteins in small and large scale experiments. This is further supporting our assumption that 1BKR and 2GKG could be as good as potential solubility tag as the Fh8 tag. (246).

Also the results of our novel tags showed that these tags maintained the correct folding and solubility of UCHL1 and UCHL3 proteins as indicated by results of the size exclusion chromatography and SAXS. The gel filtration showed that the proteins were not aggregated as they didn't elute in the peak of the void volume and eluted in the expected elution volume for each protein which correspond to the standard curve of each gel filtration column. This approach has been shown previously with the monomeric Ocr (Mocr) tag, which suggests that the novel tags can be used like Mocr tag as solubility tags (68). The SAXS results indicated correct folding of UCHL1 and UCHL3 proteins when fused with 1BKR and 2GKG tags, as no aggregation was observed by Guinier, dimensionless Kratky and Pair-Distribution Function plots, although some flexibility was observed and some of the fusions appeared more compact than others that behaved more than two-domain proteins. These results show that SAXS is a suitable method to investigate the fusion tags and may allow further studies for example using the thioredoxin tag for the structural analysis of a ternary protein-RNA complex by SAXS (247).

For crystallization, all tagged-fusions failed to form crystals except the 2GKG tag in fusion with the UCHL3 protein in one condition but the crystal did not diffract and no interpretable electron density could be observed for the target protein. The failure of

tags to aid crystallization of fusion proteins may in principle be due to the soluble tag folding correctly, but the target protein was not fully folded, which can happen (248). However, the SAXS data and gel filtration results suggested that the tags and fusion proteins had been correctly folded.

The more likely explanation is the presence of the fixed arm linker, which may create flexibility between the solubility tag and the target protein, which prevents crystal lattice formation (249). The normalized Kratky plot did suggest a degree of flexibility of the fusion proteins. UCHL1 and UCHL3 also contain flexible loop regions.

The third possible suggestion is the His tag at the N terminus, which as another flexible region may affects the formation of well-ordered diffracting crystals (250).

The construct	Yield (mg/L culture)
1BKR-UCHL1	12.45
2GKG-UCHL1	25
MBP UCHL1	27

Table 14. Comparison of yields of 1BKR-UCHL1, 2GKG- UCHL1 and MBP-UCHL1.

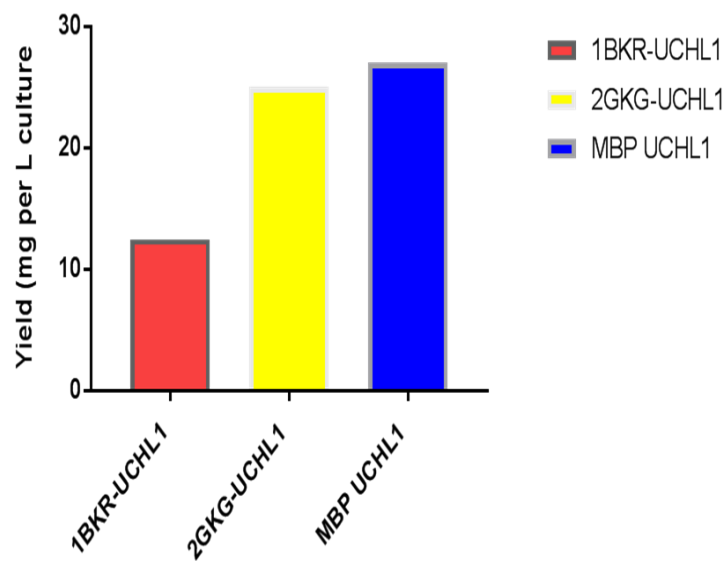


Figure 47. Graph showing a relative comparison of the yields of 1BKR-UCL1, 2GKG- UCL1 and MBP-UCL1. (no error bar because it just only one experiment).

The construct	Yield (mg/L culture)
1BKR-UCL3	6.91
2GKG-UCL3	7.6
MBP-UCL3	7.5

Table 15.Comparison of yields of 1BKR-UCL3, 2GKG- UCL3 and MBP-UCL3.

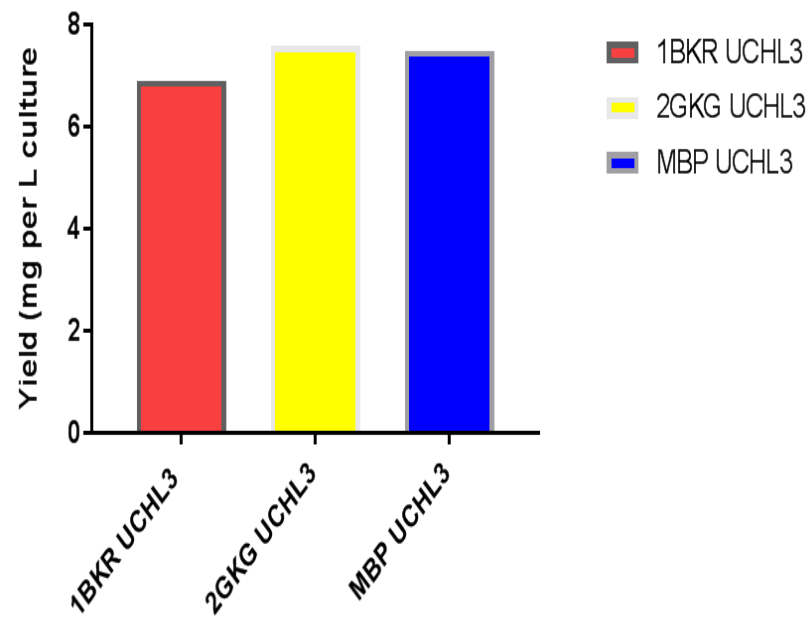


Figure 48. Graph showed the comparison of yields of 1BKR-UCHL3, 2GKG- UCHL3 and MBP-UCHL3 (no error bar because it just only one experiment).

5. Chapter 5: Results and discussion of effect of novel tags on the expression and purification of ataxin-3 (278-291) and alpha synuclein (58-79) peptides

This chapter investigated if the presence of novel tags can help improve the expression of peptides. Expression of peptides in *E coli* has been shown to be a challenge, due to peptides having a high tendency to degrade during expression by bacteria proteases (251) and this was a motivation for this research. The second aim of this chapter was to compare the two novel protein tags 1BKR and 2GKG tags and MBP tag with two model peptides.

Two peptides were chosen as their structures had been resolved previously when fused with a MBP tag making these tagged peptides a suitable control for the comparison with the novel tags. The chosen two peptides for investigation were cloned with a fixed arm linker with histidine tag in the N terminus as an affinity tag. The presence of this tag could therefore promote the expression, purification and crystallization of the fused peptides more than any other affinity tags (252). The two peptides chosen were Ataxin-3 (278-291 residues) (253) and alpha synuclein (58-79 residues) (96).

The investigation and comparison will be in terms of expression, solubility and purification in a small scale using qualitative SDS-PAGE, which was done in auto-induction conditions and in a large scale 2 litres bacterial culture following IPTG induction for quantitative comparison. Assays using at a large scale would again

facilitate production at a high-yields of soluble proteins for subsequent crystallisation trials for each model protein (254). Nickel affinity chromatography purified the proteins as a first step purification followed by gel filtration chromatography.

5.1 Results and investigation of Ataxin-3 (278-291) peptide - tag fusions

This section shows the results of the cloning, expressions, purifications and the crystallization trials of the Ataxin (278-291) peptide when fused with 1BKR, 2GKG and MBP tags using a pRSF13 vector.

5.1.1 Cloning of 1BKR-Ataxin (278-291), 2GKG-Ataxin (278-291) and MBP-Ataxin (278-291) constructs

The 1BKR-Ataxin (278-291) and 2GKG-Ataxin (278-291) constructs were cloned by annealing oligonucleotides primers protocol as mention in (material and methods). Sequencing confirmed that the 1BKR-Ataxin (278-291) and 2GKG-Ataxin (278-291) constructs were successfully cloned.

The construct of MBP-Ataxin (278-291) was engineered by a restriction enzyme free cloning protocol (material and methods) by substituting the 2GKG tag from 2GKG-ataxin (278-291) construct by MBP tag. Sequencing analysis confirmed that the MBP-Ataxin (278-291) construct had been obtained.

5.1.2 Expression, solubility and small-scale purification of 1BKR-Ataxin (278-291), 2GKG-Ataxin (278-291) and MBP-Ataxin (278-291)

This experiment was done to evaluate and compare the effect of the novel tags and MBP tag when fused with a model peptide Ataxin (278-291). The comparison in terms of expression, solubility and purification was performed in a small-scale 100 mL culture. This experiment was compared three different constructs 1BKR-Ataxin (278-291), 2GKG-Ataxin (278-291) and MBP-Ataxin (278-291).

Analysis by SDS-PAGE confirmed that all the proteins had been expressed and had an extra band in the post-induction fractions, when compared with non-induced fractions. All post-induction bands for each protein were of the correct size. The size of MBP-Ataxin (278-291) was 42 kDa, 2GKG-Ataxin (278-291) was 15.9 kDa and 1BKR-Ataxin (278-291) was 15 kDa. The bands of MBP-Ataxin (278-291) and 2GKG-ataxin (278-291) after induction were more intense than that observed for 1BKR-Ataxin (278-291) (Figures 50 A).

Analysis of the soluble fractions by SDS-PAGE gel result revealed that the soluble fractions of MBP-Ataxin (278-291) and 2GKG-Ataxin (278-291) were more intense than 1BKR-Ataxin (278-291) (Figures 50 B). SDS-PAGE analysis of the purification confirmed that the band of the MBP tag and 2GKG tag was more prominent than the 1BKR tag (Figures 50 C)

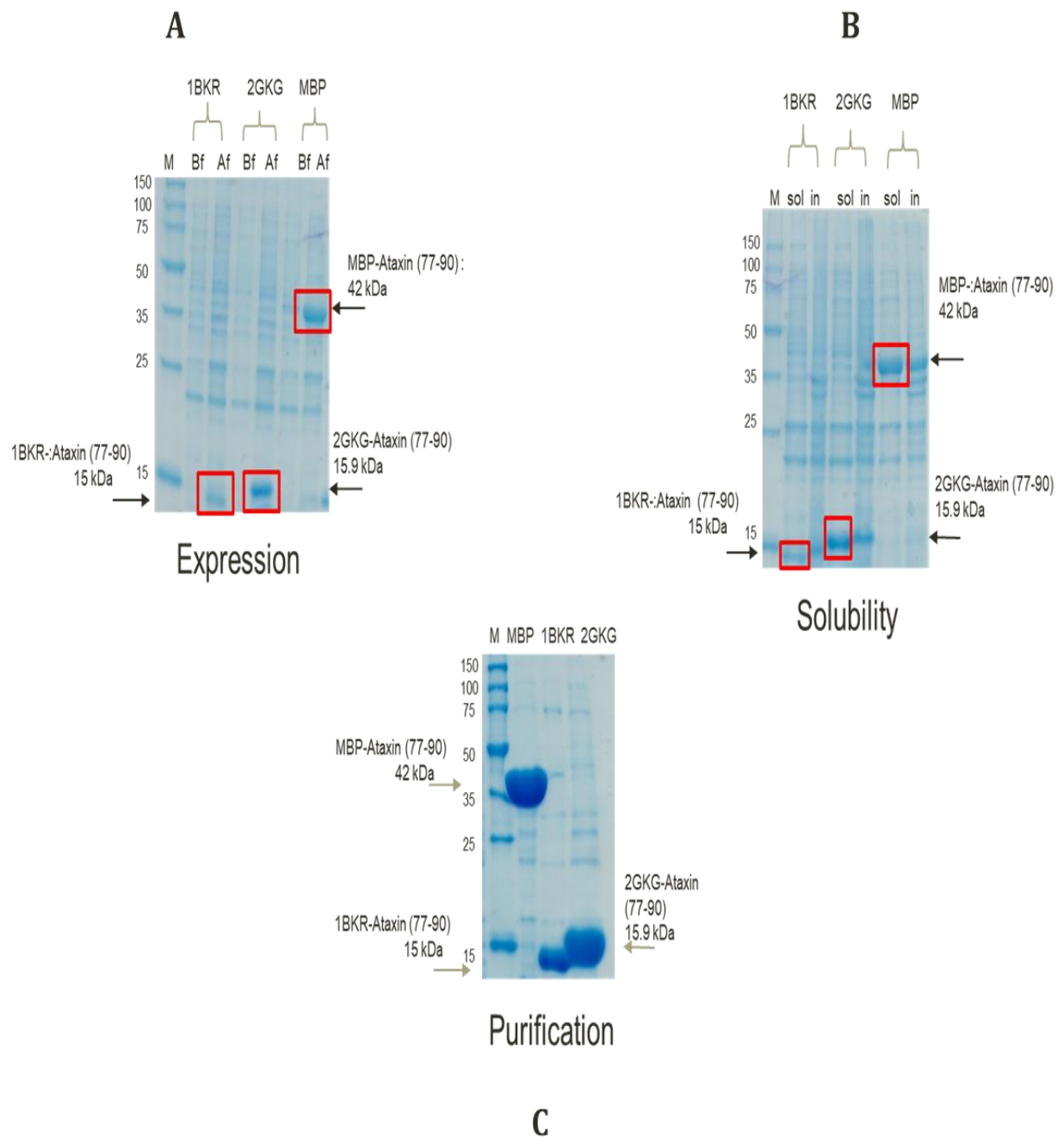


Figure 49. A: The SDS-PAGE gel of expression test indicated all the proteins were expressed and have extra band in the post-induction fractions. B: The SDS-PAGE gel of solubility test of MBP-Ataxin (278-291), 2GKG-Ataxin (278-291) and 1BKR-Ataxin (278-291). C: The SDS-PAGE gel of purification test of the three proteins MBP-Ataxin (278-291), 2GKG-Ataxin (278-291) and 1BKR-Ataxin (278-291).

5.1.3 Large-scale expression and purification of Ataxin (278-291) peptide-solubility tag fusions

5.1.3.1.1 1BKR-Ataxin (278-291) purification

The purification result of 1BKR-Ataxin (278-291) peptide by Nickel affinity chromatography indicated the protein had been well expressed and purified. The Nickel column purification of 1BKR-Ataxin (278-291) peptide showed a broad single peak, indicating a protein with high yield was obtained (Figure 51 A). The peak fractions were run on SDS-PAGE gel contained the target protein 1BKR-ataxin (278-291) characterised by an intense band at around 15 kDa, which was analogous to the protein target (Figure 51 B).

The gel filtration purification of 1BKR-Ataxin (278-291) showed that the protein eluted as a single peak which eluted at volume 79 mL (Figure 51 C). This elution volume correlates to the volume of a standard curve, for a protein of around 15 kDa in size. The gel filtration peak was very prominent, indicating a high protein yield. Analysis by SDS-PAGE gel showed that the protein purity and the bands of the 1BKR protein with and without fusion Ataxin (278-291) peptide (Figure 51 D). The total pure protein was 14 mg/L culture, without any aggregation being observed indicating that the 1BKR tag had been expressed successfully. The protein yield was suitable for crystallization trials (255).

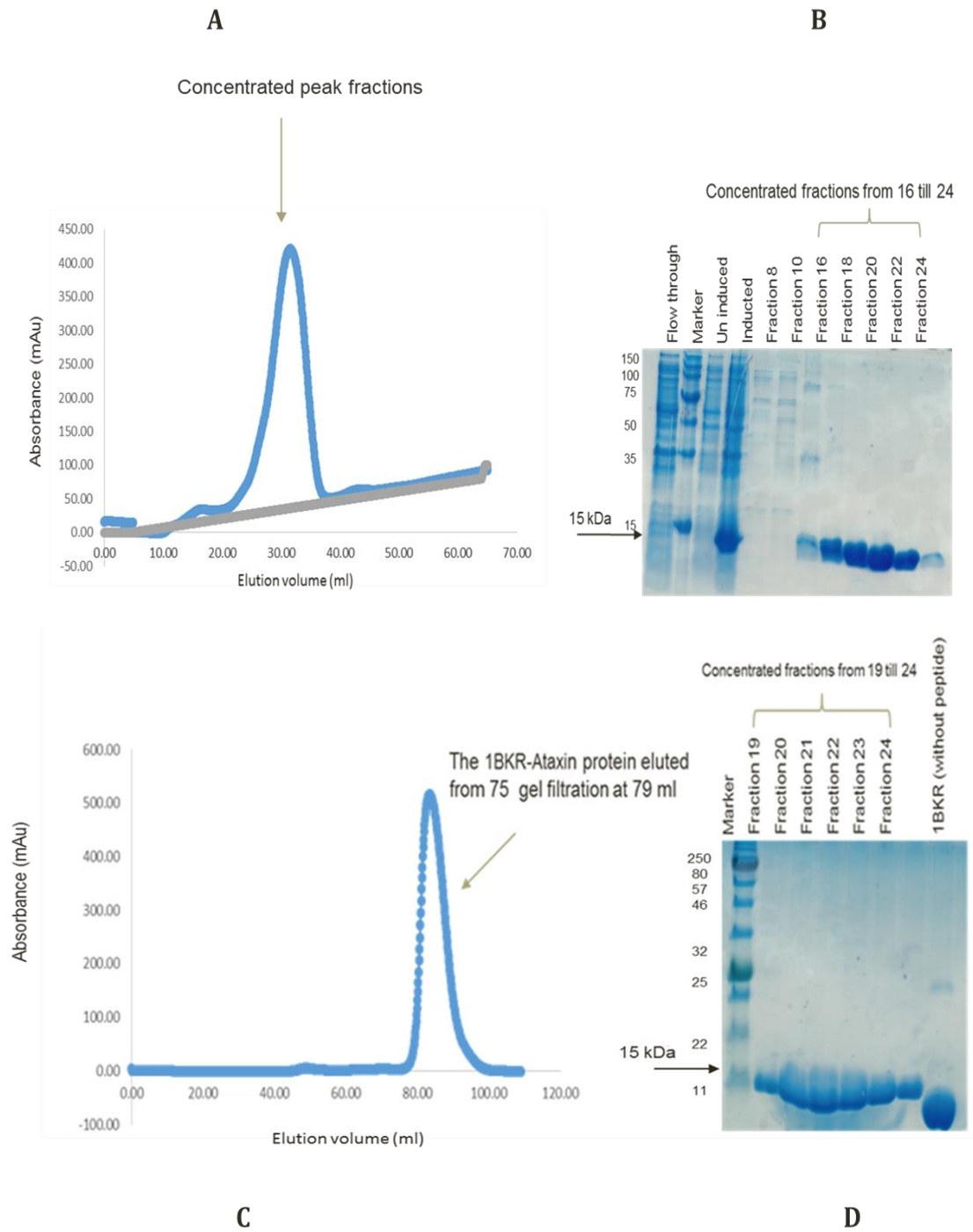


Figure 50. A: Nickel purification chromatogram of 1BKR-Ataxin (278-291) protein. B: SDS-PAGE gel from nickel column purification showed the pure protein. C: Gel filtration purification chromatogram of 1BKR-Ataxin (278-291) protein. D: SDS-PAGE gel showed purer band compared to Ni column purification with the correct size of molecular weight refer to protein marker.

5.1.3.1.2 2GKG-Ataxin (278-291) purification

The expression and purification of 2GKG-Ataxin (278-291) was performed using the same method as employed for 1BKR Ataxin (278-291) by affinity purification which showed a broad peak (Figure 52 A). Analysis of the peak fractions on a SDS-PAGE gel confirmed that the protein of interest which had been expressed and purified was 2GKG- Ataxin (278-291) (Figure 52 B). The peak fractions were concentrated and purified using gel filtration as second purification step and for protein stability assessment. The gel filtration purification revealed one sharp peak (Figure 52 C). This peak contained the protein of interest which eluted at 77 mL. This value was expected for a peptide of 16 kDa in molecular weight when using a standard curve. The SDS-PAGE gel confirmed the gel filtration chromatogram result showing an expressed protein of the correct size of approximately 15 kDa (Figure 52 D). The final yield was 22.5 mg/L cultures, which is higher than 1BKR tag yield with Ataxin (278-291) peptide, which suggests the 2GKG had a higher effect than 1BKR with Ataxin (278-291) peptide. The 2GKG-Ataxin (278-291) protein fractions were concentrated without any aggregation observed subsequently the concentrated protein was used for crystallizing trials.

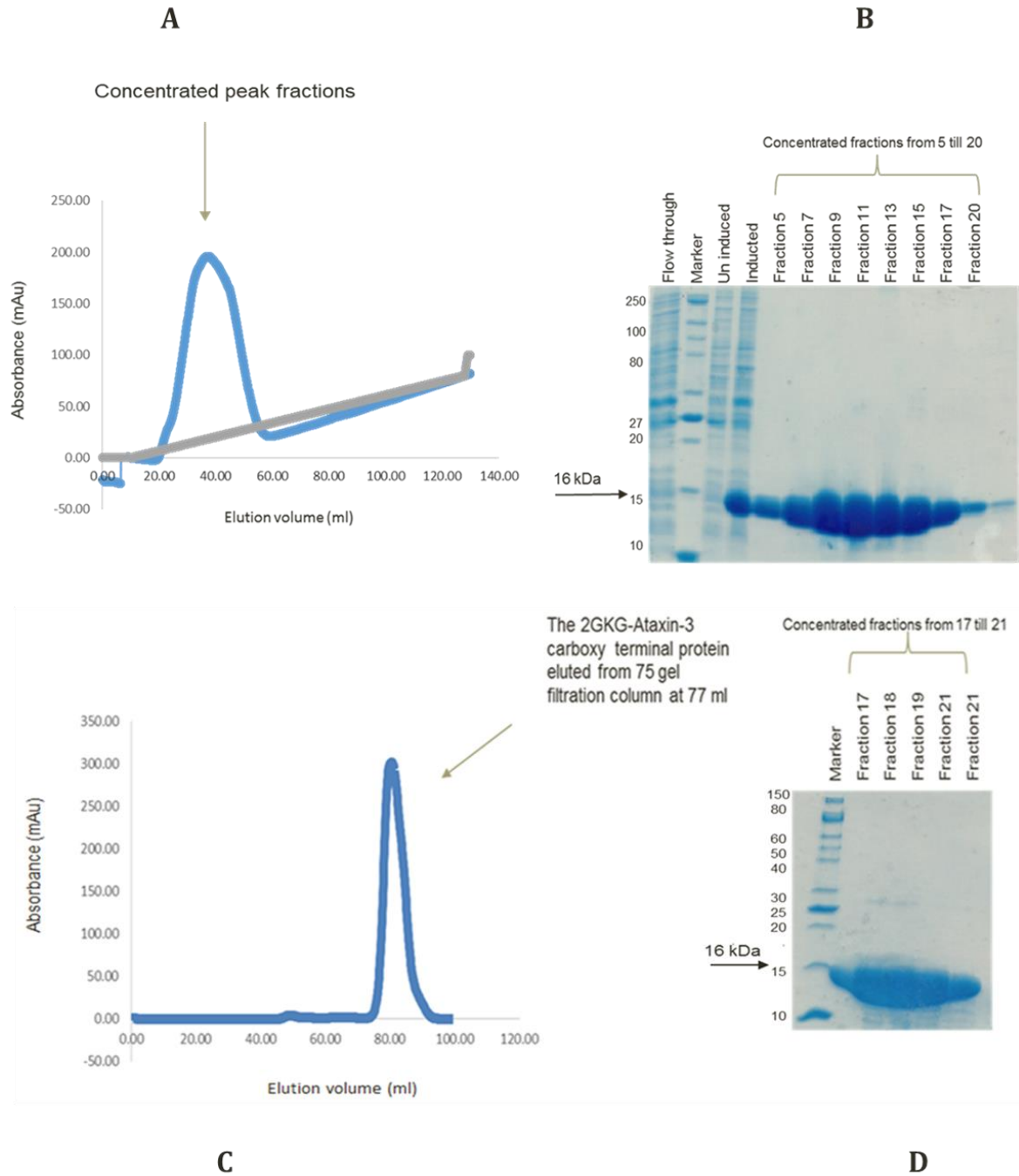


Figure 51.A: The nickel column purification chromatogram of 2GKG-Ataxin (278-291) protein. **B:** the purified fractions analysis by SDS-PAGE gel revealed the existence of a purified protein. **C:** Chromatogram after gel filtration purification of 2GKG-Ataxin (278-291) protein. **D:** SDS-PAGE gel of 2GKG-Ataxin (278-291) protein after gel filtration purification.

5.1.3.1.3 MBP-Ataxin (278-291) purification

The Nickel affinity purification of MBP-Ataxin (278-291) showed one peak on the chromatogram (Figure 53 A). SDS PAGE gel analysis revealed that the protein was well expressed, soluble and present as a clear single band of about 35 kDa in size (Figure 53 B). The predicted size of MBP-Ataxin (278-291) was 42 kDa. The chromatogram of the gel filtration purification showed one peak, the protein eluted at 58 mL (Figure 53 C), when compared to the standard curve, this volume corresponds to a protein of approximately 42 kDa. The SDS PAGE gel data indicated that the MBP-Ataxin (278-291) protein was pure as shown by a clear single band of around 42 kDa (Figure 53 D). The final yield was 23 mg/L cultures without any aggregation observed. The final soluble yield was larger than the yield of 2GKG and 1BKR tags, which suggest the MBP tag had a higher effect than the two novel tags 1BKR and 2GKG on expression of Ataxin (278-291) peptide.

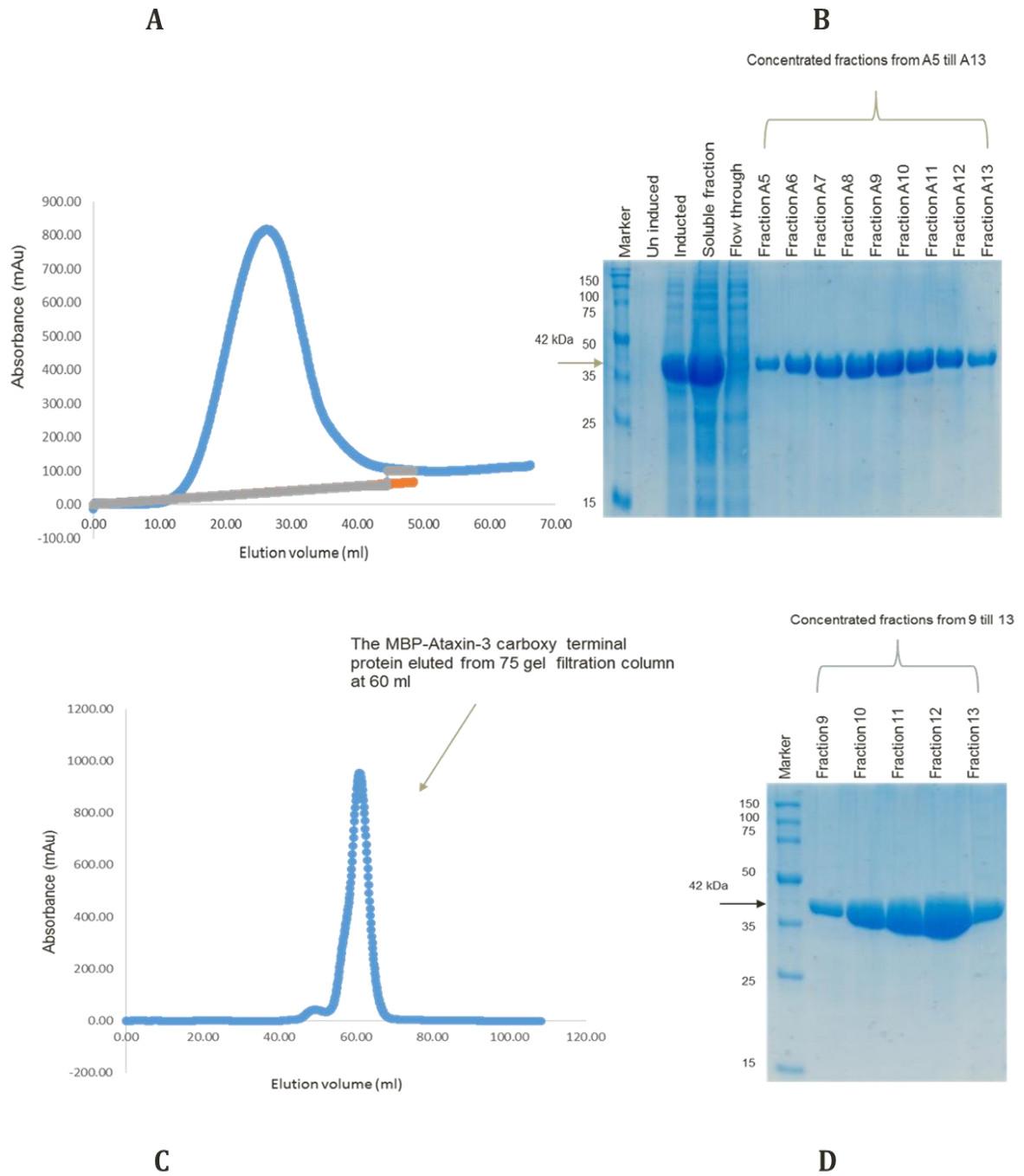


Figure 52: A: The nickel column chromatogram of MBP-Ataxin (278-291) protein. B: SDS-PAGE gel from nickel column purification of MBP-Ataxin (278-291) protein. C: Gel filtration purification chromatogram of MBP-Ataxin (278-291) protein. D: SDS-PAGE gel of MBP-Ataxin (278-291) protein after gel filtration purification.

5.1.4 Crystallisation trials of Ataxin (278-291) peptide- solubility tag fusions

Crystallisation trials of 1BKR-Ataxin (278-291), 2GKG-Ataxin (278-291) and MBP-Ataxin (278-291) proteins were performed using screens such as Morpheus screen, JCSG+, PACT premier screen, PEG Screen (Grid screen) and Ammonium sulphate (Grid screen). After one-month incubation at 20°C, a small crystal was observed in JCSG plus screen for 1BKR-Ataxin (278-291) (Figure 54). It was sent to the synchrotron for evaluation but no diffraction was obtained. The concentrations used were 10 and 20 mg/mL of the 1BKR-Ataxin (278-291) and MBP-Ataxin (278-291) proteins. Attempts were made to crystallize 2GKG-Ataxin (278-291) using 7.5,15 and 30 mg/mL, as increasing protein concentration may increase the probability to crystallize the proteins, bringing the protein up to a super saturation state (256). Using a PEG screen (grid screen) a crystal was observed that was very similar to the one observed for 1BKR-Ataxin (278-291) but it didn't diffract as well (Figure 55). The crystal screens results after one week are shown in the tables 16,17,18.



Figure 53. Potential crystals of the 1BKR-Ataxin (278-291) protein. The condition was PEG 400 40% Citric acid Buffer, pH 6.5.

Crystallization trials :1BKR-Ataxin (278-291) protein 25 mg/ml

screens	Clear	Light precipitate	Heavy precipitate
Ammonium sulphate	12	55	29
JCSG plus	14	50	32
PACT Premier	30	29	37
Morpheus	5	35	56
PEG Screen	32	40	24

Table 16. Results of crystallisations using different crystal screens for 20 mg/mL 1BKR-Ataxin (278-291) protein after one week.



Figure 54. The crystal-like objects from 2GKG-Ataxin (278-291) protein.

Crystallization trials :2GKG-Ataxin (278-291) protein 25 mg/ml

screens	Clear	Light precipitate	Heavy precipitate
AMM sulphate	24	68	4
JCSG plus	34	45	17
PACT Premier	25	36	35
Morpheus	16	32	48
PEG Screen	12	49	35

Table 17. Results of crystallisations using different crystal screens for 25 mg/mL 2GKG-Ataxin (278-291) protein after one week.

Crystallization trials :MBP-Ataxin (278-291) protein 25 mg/ml

screens	Clear	Light precipitate	Heavy precipitate
AMM sulphate	15	49	32
JCSG plus	18	44	34
PACT Premier	35	39	22
Morpheus	9	29	58
PEG Screen	21	49	26

Table 18. Results of crystallisations using different crystal screens for 20 mg/mL MBP-Ataxin (278-291) protein after one week.

5.2 Results and investigation of α -synuclein (58-79) peptide-tag fusions

This section shows the results and comparisons of second model peptide α -synuclein (58-79) peptide when fused with 1BKR, 2GKG and MBP tags in terms of expressions, purification and crystallization trials.

5.2.1 Cloning of 1BKR-alpha synuclein (58-79), 2GKG-alpha synuclein (58-79) and MBP-alpha synuclein (58-79) constructs

The alpha synuclein (58-79) gene was cloned after Histidine and 1BKR tags at the N terminus in the multiple cloning sites. Also, it was cloned in another vector after histidine and 2GKG tags at the N terminus. Sequencing revealed that these constructs had been engineered correctly.

The construct of MBP- alpha synuclein (58-79) was engineered using a restriction enzyme free cloning protocol (material and methods) substituting the 2GKG gene from the 2GKG-alpha synuclein (58-79)) construct. Sequencing analysis confirmed that a MBP-alpha synuclein (58-79) construct had been obtained.

5.2.2 Expression, solubility and small-scale purification of 1BKR-alpha synuclein, 2GKG-alpha synuclein and MBP-alpha synuclein

This experiment investigated and compared the effect of the designed novel tags with MBP tag by auto induction in a small culture scale. This experiment was compared three different constructs 1BKR-alpha synuclein (58-79), 2GKG-alpha synuclein (58-79) and MBP-alpha synuclein (58-79).

Analysis by SDS-PAGE of the un-induced and induced fractions of all the proteins indicated all the proteins had been successfully expressed being characterised by the presence of extra protein band in the post-induced sample (Figure 56 A). All the post-induction bands for each protein were of the correct size which were MBP-alpha synuclein (58-79) (42 kDa), 2GKG- alpha synuclein (58-79) (15.9 kDa) and 1BKR-alpha synuclein (58-79) (15 kDa).

The post induction expression band of both MBP-alpha synuclein (58-79) and 2GKG-alpha synuclein (58-79) were almost the same. Both being more pronounced than the post induction expression band of 1BKR-alpha synuclein (58-79). This suggested that MBP and 2GKG tags had the higher effect than 1BKR tag on expression of alpha synuclein (58-79) peptide.

Analysis on SDS-PAGE gel of the soluble protein fractions of the three proteins 1BKR-alpha synuclein (58-79), 2GKG-alpha synuclein (58-79) and MBP-alpha synuclein (58-79) revealed that the MBP-alpha synuclein (58-79) soluble band was

more pronounced than that observed for 2GKG- α synuclein (58-79) and 1BKR- α synuclein (58-79) (Figure 56 B).

The SDS-PAGE gel indicated that the MBP tagged protein band after purification was more intense than the bands observed for the other tags (Figure 56 C).

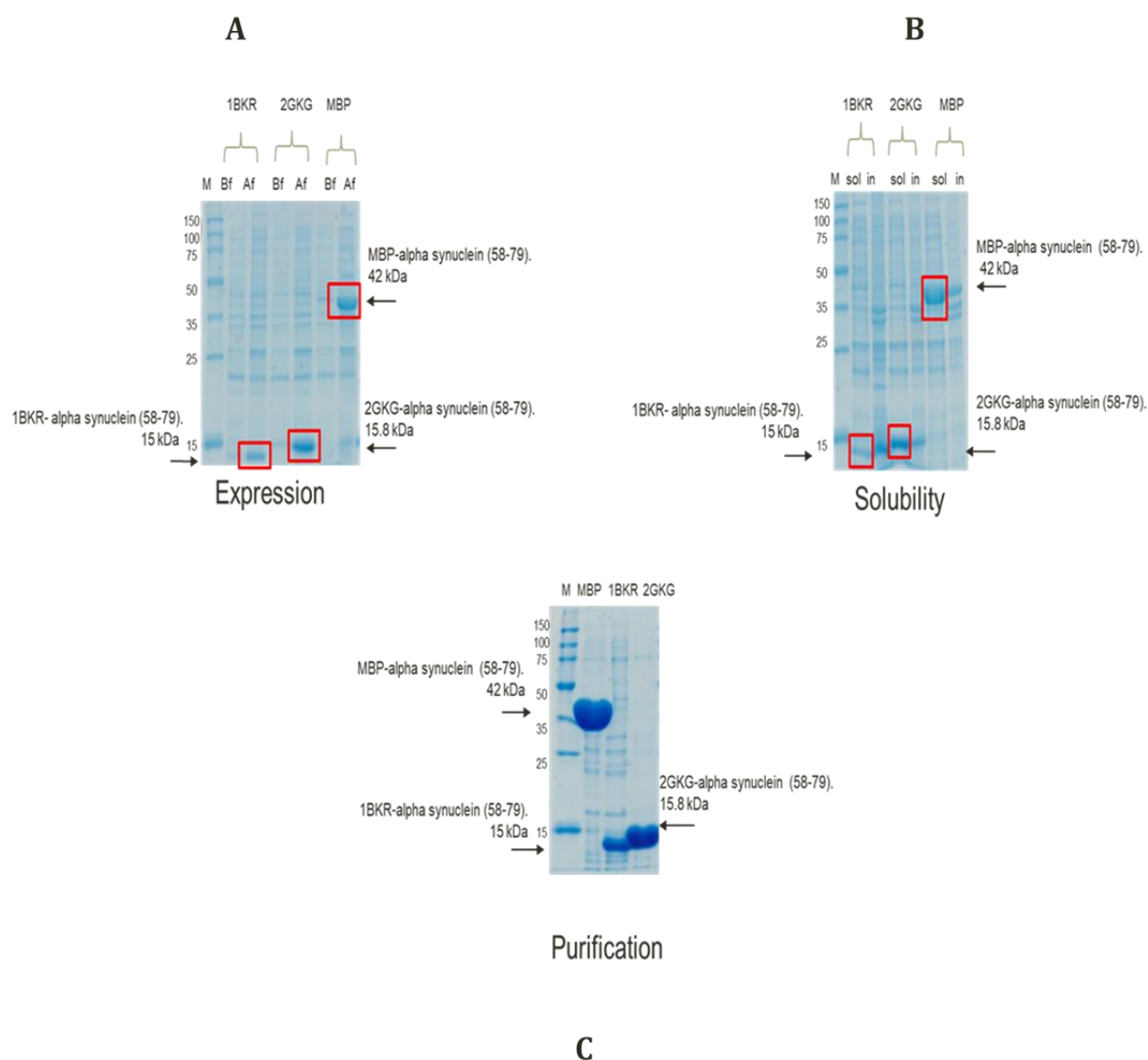


Figure 55. A: The SDS-PAGE gel of expression test of the three proteins 1BKR- α synuclein (58-79), 2GKG- α synuclein (58-79) and MBP with α synuclein (58-79). B: The SDS-PAGE gel of solubility test of the three proteins 1BKR- α synuclein (58-79), 2GKG- α synuclein (58-79) and MBP with α synuclein (58-79). C: The SDS-PAGE gel of purification test of the three proteins MBP- α synuclein (58-79) 2GKG- α synuclein (58-79) and 1BKR- α synuclein (58-79).

5.2.3 Large-scale expression and purification of alpha synuclein (58-79) peptide-solubility tag fusions

5.2.3.1 1BKR- alpha synuclein (58-79) protein purification

Nickel affinity chromatogram of the 1BKR-alpha synuclein (58-79) revealed a prominent peak (Figure 57 A). All the peak fractions were run on a SDS-PAGE gel which indicated that the protein had been successfully expressed, bound to the nickel column and purified with some contaminants (Figure 57 B).

The chromatogram of gel filtration step revealed an elution volume of protein at 77 mL (Figure 57 C) which correlated with a protein of 15 kDa, based on the standard curve. Peak fractions analysis by SDS-PAGE gel showed a band which correlated to the predicted size of 1BKR alpha synuclein (58-79) protein (Figure 57 D). The peak fractions were concentrated for crystallization trials and the total protein yield was 5.38 mg/L.

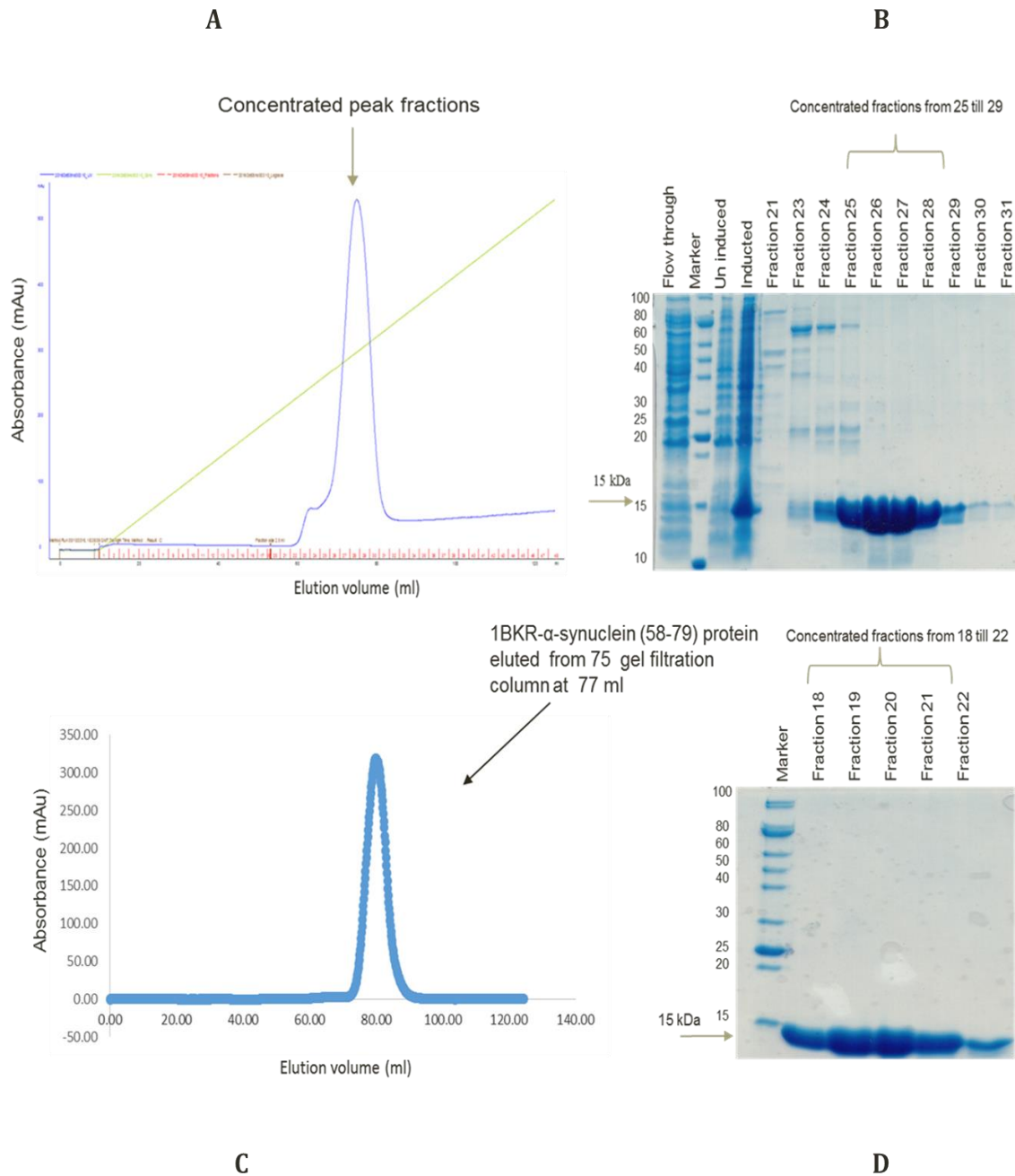


Figure 56. A: Nickel column chromatogram of 1BKR-alpha synuclein (58-79). B: The SDS-PAGE gel of purified 1BKR-alpha synuclein (58-79) samples. C: Gel filtration purification chromatogram of 1BKR-alpha synuclein (58-79). D: 1BKR-alpha synuclein (58-79) was expected on 15 kDa as shown on the SDS-PAGE gel.

5.2.3.2 2GKG-alpha synuclein (58-79) protein purification

The 2GKG-alpha synuclein (58-79) has a molecular weight of 15.8 kDa and purification was performed in the same manner as with 1BKR-alpha synuclein (58-79) protein. Nickel affinity chromatogram of the 2GKG-alpha synuclein (58-79) showed a single peak (Figure 58 A). All the peak fractions were run on a SDS-PAGE gel, which indicated the protein had been purified with little contamination (Figure 58 B). The 2GKG-alpha synuclein (58-79) protein bands appeared at approximately 15 kDa (Figure 58 B).

Gel filtration purification of 2GKG-alpha synuclein (58-79) showed two peaks, the first peak corresponded to the void volume and the second-high peak was the protein target, which eluted at 76 mL (Figure 58 C). This elution volume correlated with a protein of 15.8 kDa when compared to a standard curve. Analysis of the fractions by SDS-PAGE gel revealed a band which correlated with a protein of molecular weight around 15 kDa (Figure 58 D). SDS-PAGE gel confirmed that the protein was expressed and purified in high yield. The peak fractions were concentrated for crystallization trials. The protein yield of this purification was 15 mg/L culture without any aggregation observed which suggests that the presence of a 2GKG tag had more effect than 1BKR tag with alpha synuclein (58-79) peptide, in terms of expression and purification.

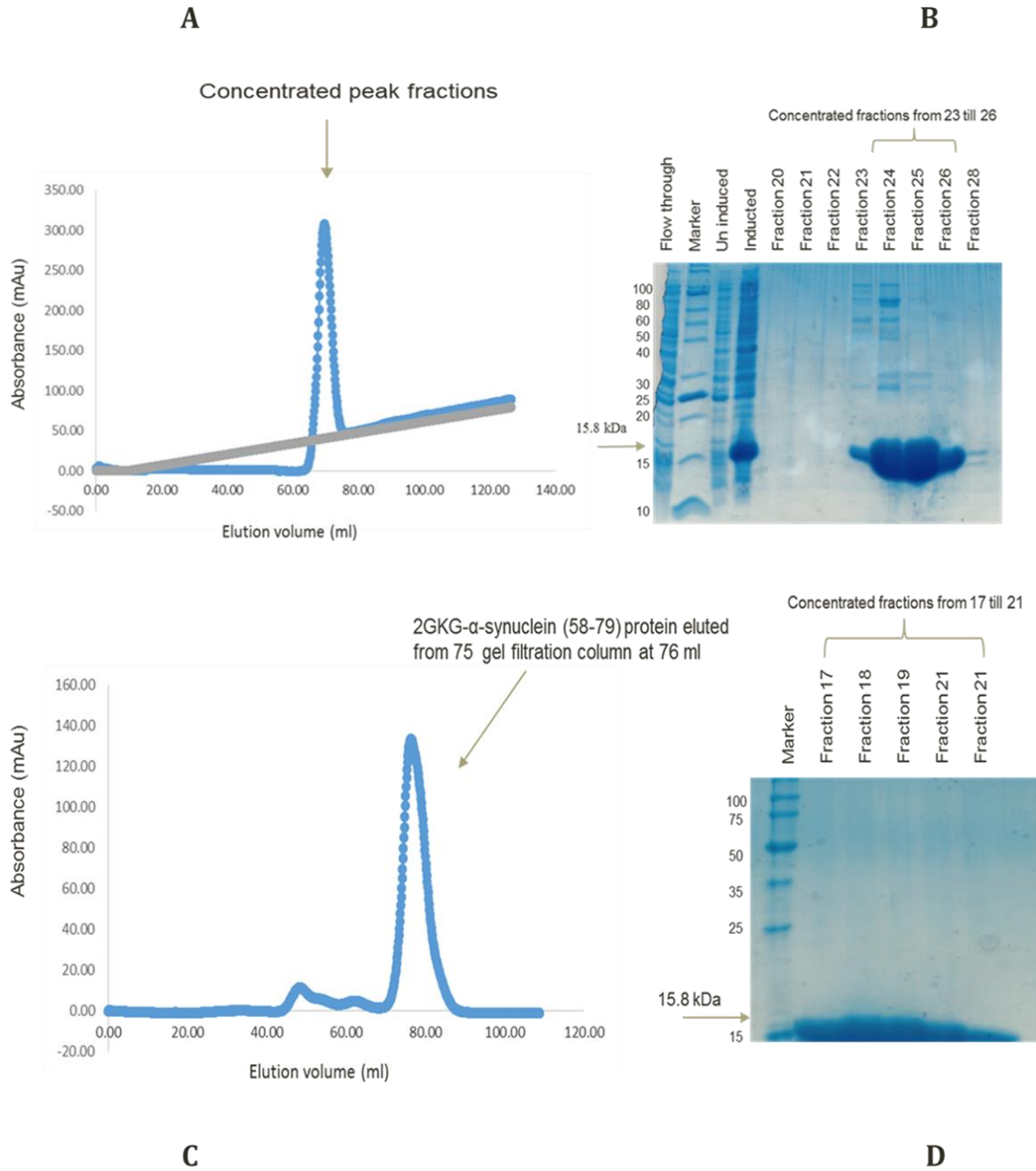


Figure 57. A: Nickel affinity purification chromatogram of 2GKG- α synuclein (58-79). **B:** The SDS-PAGE gel indicates the expression and purification of the 2GKG α synuclein (58-79) of the predict size 15.8 kDa. **C:** Gel filtration purification chromatogram of 2GKG- α synuclein. **D:** 2GKG α synuclein (58-79) was expected on 15.8 kDa as shown on the SDS-PAGE gel.

5.2.3.3 MBP-alpha synuclein (58-79) protein purification

Purification of MBP-alpha synuclein (58-79) protein used the same methods as mentioned earlier with 2GKG-alpha synuclein (58-79) and 1BKR-alpha synuclein (58-79) proteins. The Nickel column purification did not work properly as the column had some air bubbles, so no protein chromatogram was observed. All the fractions were run on SDS-PAGE gel for which showed that the MBP-alpha synuclein (58-79) protein had been expressed and purified (Figure 59A).

The gel filtration chromatogram of the purified proteins showed a high quantity product (Figure 59 B). There was a single high peak which contained the target protein eluted at 60 mL. This elution volume refers to 42 kDa, based on the standard curve, showed that MBP alpha synuclein was in a monomer formation. SDS-PAGE gel represented a high quantity of protein product at the right molecular weight (42 kDa) (Figure 59 C). The protein yield after gel filtration purification of the MBP-alpha synuclein (58-79) protein was 22.18 mg/L culture, without any aggregation observed during concentration. This suggested that the MBP tag has the highest effect when compared with the expression and purification profiles obtained with the 1BKR and 2GKG tags with alpha synuclein (58-79).

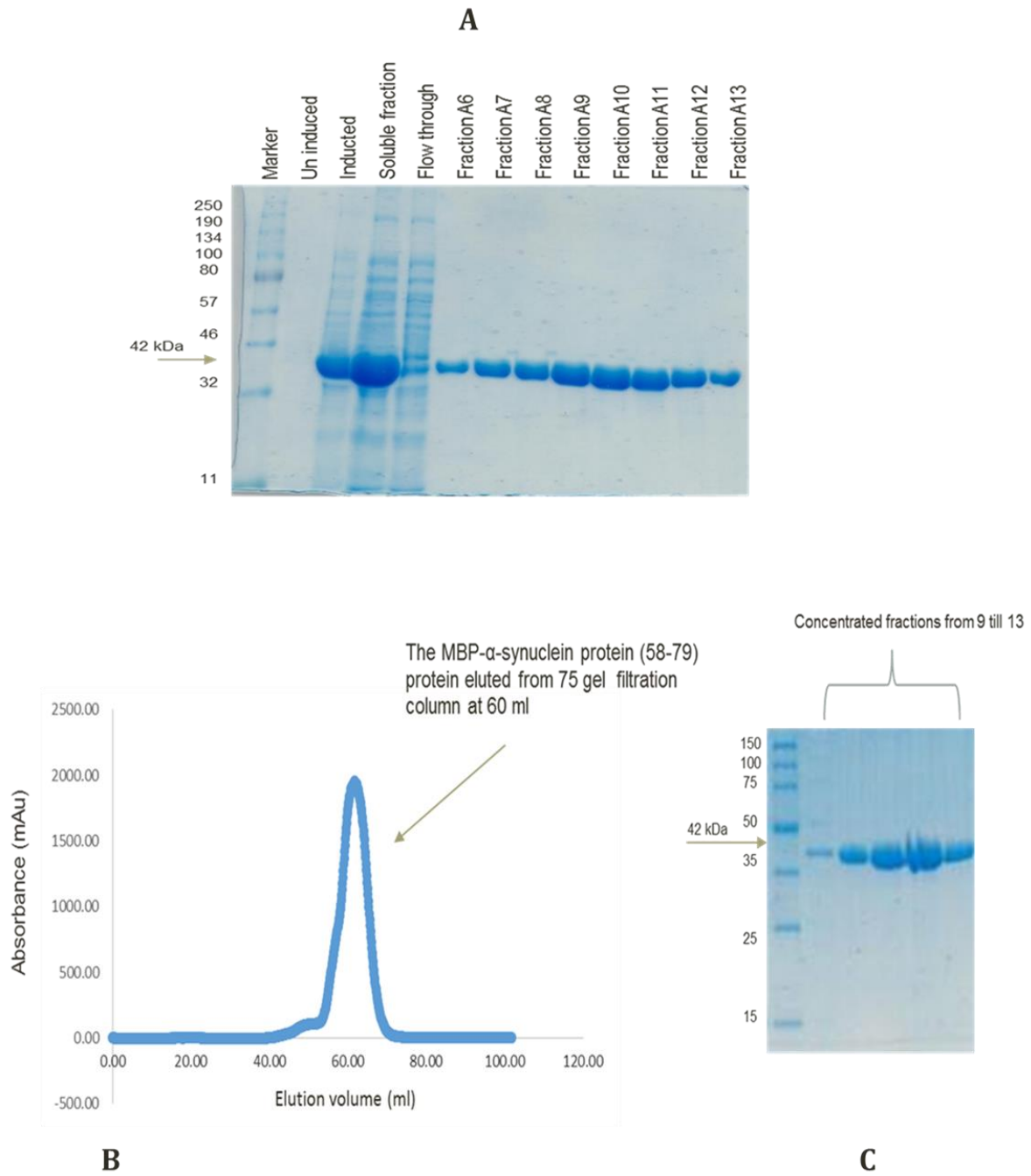


Figure 58. A: The chromatogram of affinity chromatography results of MBP- α synuclein (58-79). B: SDS-PAGE gel result C: Nickel column purification Chromatogram of MBP- α synuclein (58-79). D: SDS-PAGE gel result showed bands that correspondent to MBP- α synuclein (58-79) (42 kDa).

5.2.4 Crystallization trials of alpha synuclein (58-79) peptide-solubility tag fusions

Crystallisation trials were carried out on 1BKR-alpha synuclein (58-79), 2GKG-alpha synuclein (58-79) and MBP- alpha synuclein proteins using different screens such as Morpheus, JCSG+, PACT premier, PEG (Grid screen) and Ammonium sulphate (Grid screen). The incubation temperature of the attempts was 20°C. The concentrations of the proteins used were 7.5 and 15 mg/mL. The 1BKR-alpha synuclein (58-79) yielded some light precipitates that were observed in most of the screens after several months (Figure 60). After several weeks for 2GKG-alpha synuclein (58-79) protein a very small crystal was obtained in the PACT screen but no diffraction was obtained (Figure 61). The crystal screen results after one week are presented in table 19. Crystallisation trials with MBP alpha synuclein failed to observe any significant crystal growth. The crystal screens result after one week are shown in table 21.



Figure 59. Precipitate from 1BKR-alpha synuclein (58-79) protein.

Crystallization trials: 1BKR-alpha synuclein (58-79) protein 15 mg/mL			
screens	Clear	Light precipitate	Heavy precipitate
AMM sulphate	20	52	24
JCSG plus	15	38	43
PACT Premier	29	42	25
Morpheus	15	31	50
PEG Screen	28	34	34

Table 19. Results of crystallisation using different crystal screens for 15 mg/mL 1BKR-alpha synuclein (58-79) protein after one week.



Figure 60. The crystal from 2GKG-alpha synuclein (58-79) protein in the PACT screen (0.2 M Sodium chloride 0.1 M MES, 20 % w/v PEG 6000, pH- 6.0).

Crystallization trials: 2GKG-alpha synuclein (58-79) protein 20 mg/mL			
screens	Clear	Light precipitate	Heavy precipitate
AMM sulphate	12	58	26
JCSG plus	26	41	29
PACT Premier	37	49	9
Morpheus	11	37	48
PEG Screen	21	39	36

Table 20. Results of crystallisation using different crystal screens for 20 mg/mL 2GKG-alpha synuclein (58-79) protein after one week.

Crystallization trials: MBP alpha synuclein protein 20 mg/mL			
screens	Clear	Light precipitate	Heavy precipitate
PEG Screen	24	47	25
JCSG plus	26	27	43
PACT Premier	30	29	37
Morpheus	17	32	47
AMM sulphate	13	42	41

Table 21. Results of crystallisation using different crystal screens for 20 mg/mL MBP alpha synuclein protein after one week.

5.3 Discussion of the results of ataxin (278-291) and alpha synuclein (58-79) peptide fusions with the tags

This chapter discussed the results for 1BKR, 2GKG and MBP tags when fused with Ataxin (278-291) and Alpha synuclein (58-79) peptides. The production of these peptides in *E coli* required a fusion tag as the peptide is susceptible to degradation by bacterial proteases (257).

The results from small-scale expression trials indicated that the 1BKR and 2GKG proteins successfully protected the peptide from cellular proteases, analysis also revealed that these peptides were expressed and purified near to the level when in fusion with the MBP tag. This data supported our hypothesis that 1BKR and 2GKG could be used as solubility tags for the expression and purification of peptides as they have the ability to increase the amount of the soluble protein target to the level near to that observed for the MBP tag (258).

The large scale results confirmed the results from the small scale experiments indicating that the tags protected these peptides from cellular proteases. Presence of the tags also promoted correct folding resulted in non-void peaks on size exclusion chromatography as the proteins had eluted in monomeric forms and in the expected elution volume for each protein (259). These results revealed that the presence of the MBP and 2GKG tags had a higher effect on the expression, solubility and purification of Ataxin (278-291) and alpha synuclein (58-79) peptides, when compared with the 1BKR tag (Table 22,23 and Figures 62,63). The reason that the MBP tag fusion had a little higher yield than the 2GKG tag may be due to its larger size as the presence of a larger fusion tag can increase the rates of translation of the fusion partner (260).

For crystallization, the 1BKR and 2GKG tagged fusion Ataxin (278-291) and alpha synuclein (58-79) peptides failed to crystallize. The first suggested explanation of this failure may be due to the two novel tags having a small molecular weight which prevents the protein molecules proper packing in the crystal, not like large tags which have a large molecular weight such as glutathione-S-transferase (26kDa) or maltose-binding protein (MBP) (42kDa) tags (104) (261).

The second suggested explanation of the failure of our novel tags and MBP tag to crystallize these peptides may be the presence of the His tag at the N terminus. This suggested as the MBP tag crystalized these peptides previously without His tag as reported by Zhao et al and Zhemkov et al (96) (253). The presence of His tag at the N terminus may introduce more flexibility which may hinder the crystal formation (252).

The construct	Yield (mg per L culture)
1BKR-Ataxin	14
2GKG-Ataxin	22.5
MBP-Ataxin	23

Table 22.Comparison of yields of 1BKR-Ataxin (278-291), 2GKG-Ataxin (278-291) and MBP-Ataxin (278-291) constructs after gel filtration purification.

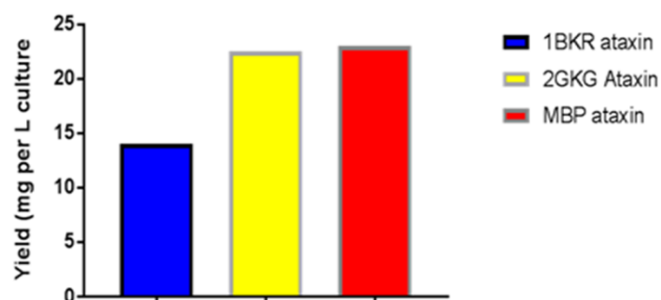


Figure 61. Graph showing a comparison of the yield of 1BKR- Ataxin (278-291), 2GKG- Ataxin (278-291) and MBP Ataxin (278-291) constructs after gel filtration purification (no error bar because it just only one experiment).

The construct	Yield (mg/L)
1BKR-alpha synuclein (58-79)	5.38
2GKG-alpha synuclein (58-79)	15
MBP-alpha synuclein(58-79)	22.18

Table 23.Comparison of yields of MBP- alpha synuclein (58-79) 2GKG- alpha synuclein (58-79) and 1BKR- alpha synuclein (58-79).

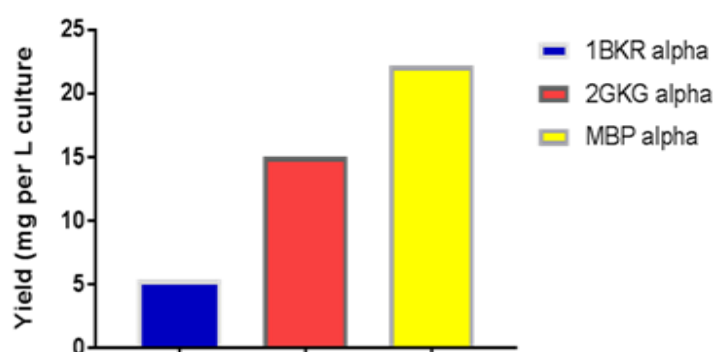


Figure 62.Graph showed the comparison of yield of MBP- alpha synuclein (58-79) 2GKG- alpha synuclein (58-79) and 1BKR alpha synuclein (58-79) (no error bar because it just only one experiment).

6. Chapter 6: Results and discussion of novel tags fused to proteins of unknown structure, USP20 ZNFUBP (1-108) and BBS5 (1-270)

This chapter focused on the investigation and comparison of novel protein tags (1BKR and 2GKG tags) with two-model proteins of unknown structure which have been shown to be difficult to express in *E coli*. These novel tags were compared to Histidine tag and maltose binding protein (MBP) tag these being amongst the most commonly used tags. The chosen model proteins were human Bardet Biedl protein 5 BBS5 residues (1-270) spanning PH domains and the ubiquitin specific protease 20 zinc finger domain USP20 ZNFUBP residues (1-108).

A comparison was carried out firstly between Histidine tag with and without novel tags. The comparison was done at a small scale using qualitative SDS-PAGE by auto-induction condition. This comparison clarified the effect of two novel tags on the two model proteins, USP20ZNFUBP (1-108) and BBS5(1-270) in terms of expression, solubility and purification.

Secondly results of maltose binding protein (MBP) tag with model proteins were compared to that of novel tags with these model proteins. The comparison was done by large scale 2 litres bacterial culture using IPTG induction for quantitative comparison. The proteins have been purified in the same way as previously discussed with other fusion constructs using Nickel affinity chromatography and then gel filtration chromatography.

6.1 Results of investigation of the first model protein USP20 ZNFUBP (1-108)-tag fusions

6.1.1 Cloning constructs of 1BKR-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and MBP-USP20ZNFUBP (1-108)

Three constructs of USP20ZNFUBP (1-108) were generated in this section which were 1BKR-USP20ZNFUBP(1-108), 2GKG-USP20ZNFUBP(1-108) and MBP-USP20 ZNFUBP (1-108). The first construct was 1BKR-USP20ZNFUBP (1-108) which was cloned with a histidine tag at the N terminus using restriction sites *Not*I and *Xho*I in the pRSF-13 DNA vector.

The second construct was 2GKG-USP20ZNFUBP (1-108) which was cloned using the pRSF-13 vector which inserted a 2GKG tag and histidine tag at the N terminus.

The MBP-USP20ZNFUBP (1-108) construct was created using restriction enzyme free cloning protocol using the 2GKG-USP20ZNFUBP (1-108) construct to engineer MBP-USP20ZNFUBP (1-108) by substituting the 2GKG tag with the MBP tag. Sequencing of the resulting MCS region of the vector confirmed that MBP- USP20 ZNFUBP (1-108) had been engineered correctly.

6.1.2 Expression, solubility and small-scale purification of 1BKR-USP20ZNFUBP (1-108), 2GKG-USP20 ZNFUBP (1-108) and His-USP20ZNFUBP (1-108)

This experiment investigated the role and influence of designed tags with USP20ZNFUBP (1-108) protein in terms of expression, solubility and purification. This experiment compared between three different constructs His-USP20 ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and 1BKR-USP20ZNFUBP (1-108) at a small scale.

Analysis on a SDS-PAGE revealed that all the proteins had been expressed as indicated by the presence of an extra protein band after induction where no bands were present in the before induction fractions (Figure 64 A). All the protein bands after induction were of the predicted size for each protein as His-USP20 ZNFUBP (1-108) (12 kDa), 2GKG-USP20ZNFUBP (1-108) (26 kDa) and 1BKR-USP20ZNFUBP (1-108) (25 kDa). The bands of 2GKG-USP20ZNFUBP (1-108) and 1BKR-USP20ZNFUBP (1-108) after induction were more intense than the protein band observed for His-USP20ZNFUBP (1-108) which indicated that these tags had improved expression of USP20 ZNFUBP (1-108) protein. The SDS-PAGE result showed the bands of the soluble fractions of 2GKG-USP20ZNFUBP (1-108) and 1BKR-USP20ZNFUBP (1-108) were more intense than that observed for His-USP20ZNFUBP (1-108) (Figure 64 B).

Analysis of the SDS-PAGE indicated the improving the solubility and the expression of USP20ZNFUBP (1-108) protein when fused to either a 1BKR or 2GKG tag. After purification of the samples using a Nickel column, analysis of a SDS-PAGE

gel revealed that the 2GKG-USP20ZNFUBP (1-108) protein band was more prominent than the protein bands observed for His-USP20ZNFUBP (1-108) and 1BKR-USP20ZNFUBP (1-108) (Figure 64 C).

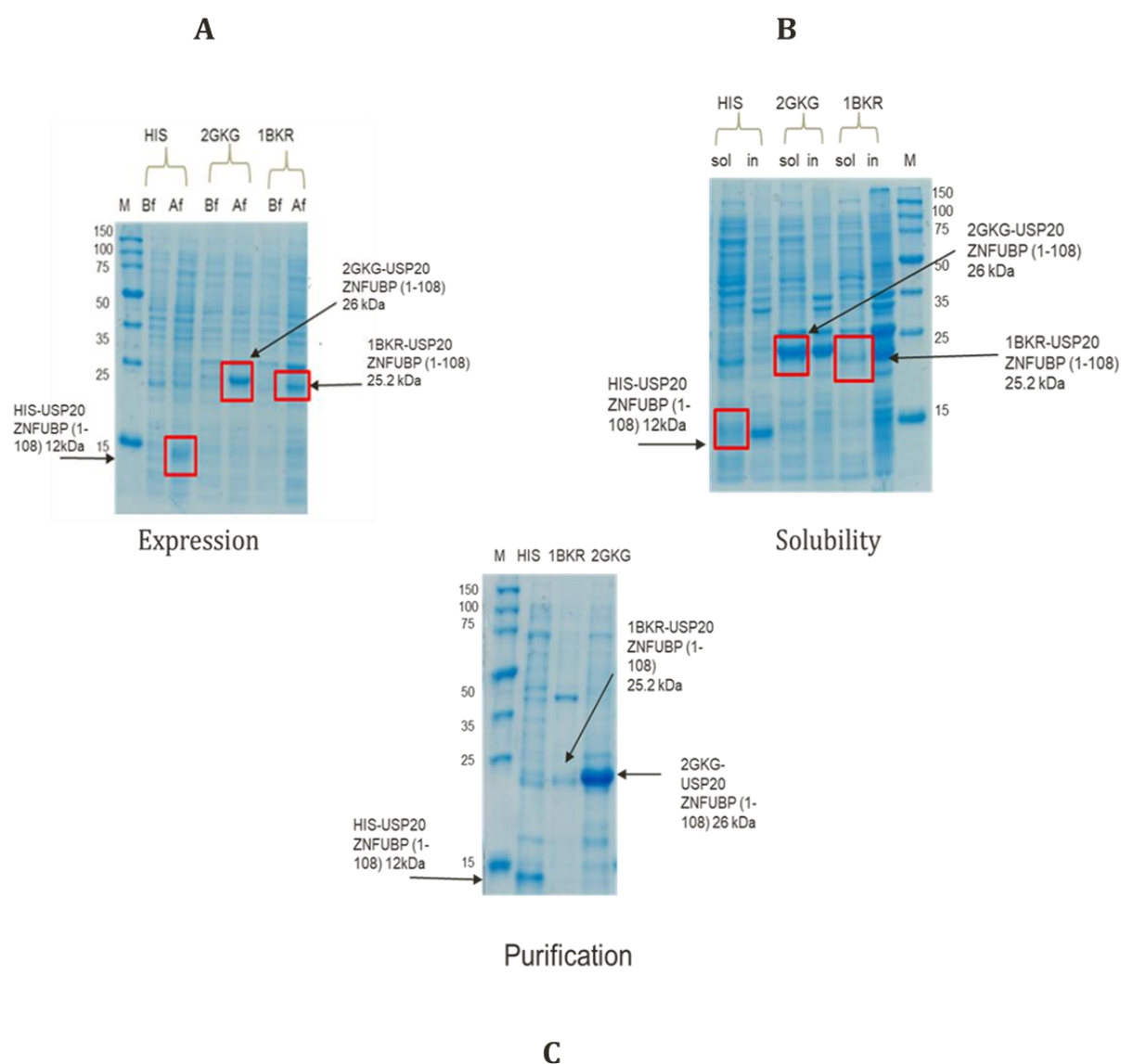


Figure 63.A: The SDS-PAGE gel of expression test of the three proteins His-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and 1BKR-USP20ZNFUBP (1-108). **B:** The SDS-PAGE gel of solubility test of the three proteins His-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and 1BKR-USP20ZNFUBP (1-108). **C:** The SDS-PAGE gel of purification test of the three proteins His-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and 1BKR-USP20ZNFUBP (1-108).

6.1.3 Large-scale expression and purification of USP20ZNFUBP (1-108) protein tag fusions

6.1.3.1 1BKR-USP20 ZNFUBP (1-108) purification

The Nickel affinity chromatography of 1BKR-USP20 ZNFUBP (1-108) showed a small peak and a broad peak (Figure.65A). The fractions under both peaks were run on SDS-PAGE gel. The first peak fractions showed no target protein. However, the second peak contained the 1BKR-USP20ZNFUBP (1-108) protein as an intense band below 27 kDa of the marker which is analogous to the predicted size of the fusion protein (Figure.65 B).

The gel filtration purification, 1BKR-USP20 ZNFUBP (1-108) showed that the protein eluted as one peak (Figure 65 C). The protein eluted at elution volume at 75 mL which corresponds to the protein being a monomer. This elution volume correlated with a protein of 25 kDa in size when compared with a standard curve. The SDS-PAGE gel also confirmed that the protein was pure with a 1.25 mg/L yield (Figure 65 D).

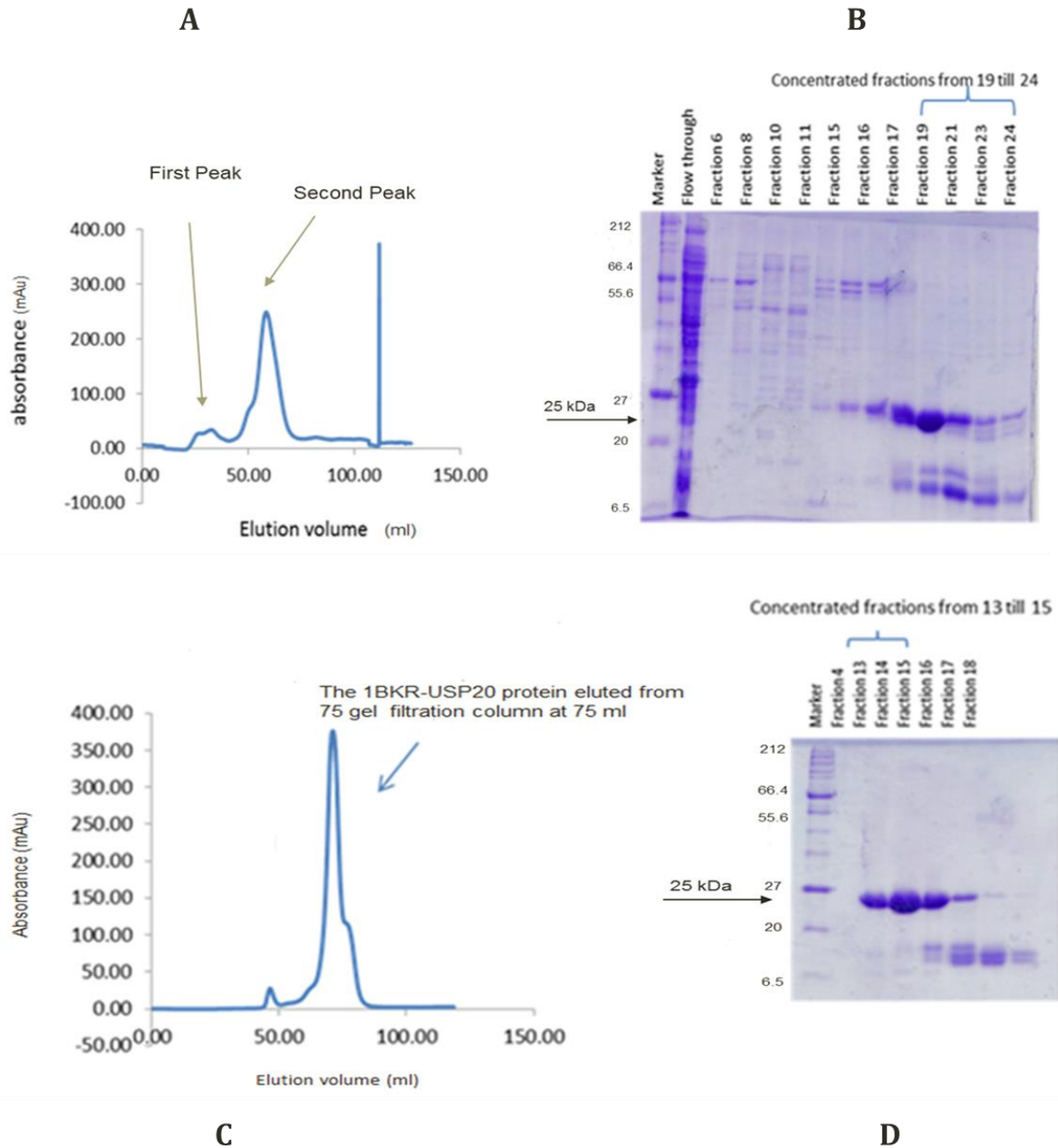


Figure 64. A: The chromatogram affinity chromatography of 1BKR-USP20ZNFUBP (1-108). B: SDS-PAGE gel from Ni column showed 1BKR-USP20 ZNFUBP (1-108) C: Chromatogram of 1BKR-USP20 ZNFUBP (1-108) gel Filtration Purification. D: SDS-PAGE gels revealed a band at around 25 kDa that is a 1BKR-USP20 ZNFUBP (1-108).

6.1.3.2 2GKG-USP20ZNFUBP (1-108) purification

The expression and purification of 2GKG-USP20 ZNFUBP was done utilizing the same way as 1BKR-USP20ZNFUBP (1-108). The purification using Nickel chromatography, two peaks were observed (Figure 66 A). These two peaks fractions were analysed on SDS PAGE gel and it was confirmed that the two peaks contained the 2GKG-USP20ZNFUBP (1-108) protein (Figure 66 B). Each peak was gel filtered separately by two different purification process using a gel filtration column (Superdex 75 HiLoad 16/60).

The first gel filtration run revealed two peaks. The first peak was the void volume peak at 45.5 mL and the second peak was 2GKG-USP20ZNFUBP with a molecular weight of 26 kDa and eluted between 65.9 ml and 77.1 mL (Figure 66 C). This value was equal to a protein with a molecular weight of 26 kDa when compared with a standard curve. The result of the second run on gel filtration column showed only one peak in the void volume (Figure 66 D). The fractions from the both gel filtration runs were analysed by SDS PAGE gel (Figure 66 E). Only the fractions from the first purification process were concentrated for crystallisation trials. The final yield was 0.86 mg/L cultures. This amount was less than that observed with a 1BKR tag, which suggest the 1BKR tag improved the stability and expression more than that observed with a 2GKG tag.

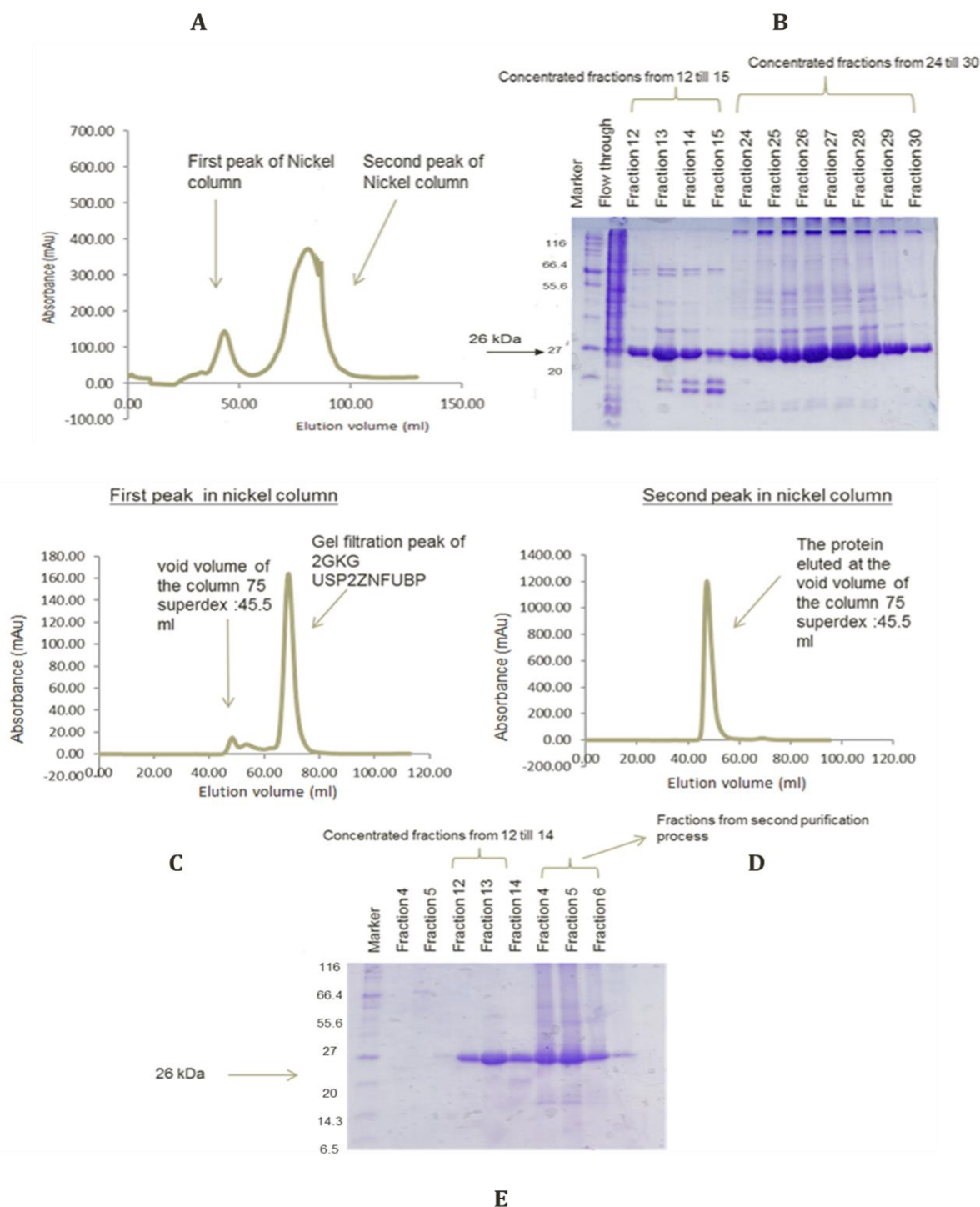


Figure 65: A: Chromatogram of affinity chromatography of 2GKG USP20 ZNFUBP (1-108). B: SDS-PAGE gel from Ni column showed the two peaks contained the 2GKG- USP20ZNFUBP (1-108) protein C: first gel filtration purification chromatogram of 2GKG-USP20ZNFUBP (1-108) from first peak of nickel purification. D: Second gel filtration purification chromatogram of 2GKG-USP20ZNFUBP (1-108) from second peak of nickel purification. E: SDS-PAGE gels showed the two gel filtration purifications only the fractions second peak were collected for crystallization trials.

6.1.3.3 MBP-USP20ZNFUBP (1-108) purification

The Nickel affinity column purification of MBP-USP20ZNFUBP (1-108) showed a peak on the chromatogram (Figure 67 A). Further analysis of purified proteins on SDS-PAGE confirmed the presence of the target fusion protein as distinct single band at about 53 kDa (Figure 67 B).

Analysis of the chromatogram of gel filtration purification revealed two peaks (Figure 67 C). The first peak could be the dimer but not the void oligomer. The second peak contained the protein target as monomer which eluted at 87.8 mL. Comparing to the standard curve, this volume of eluate refers to a protein of about 53 kDa. The SDS PAGE gel data indicated that the MBP-USP20ZNFUBP (1-108) protein was purified as a clear single band at about 53 kDa (Figure 67 D). After collection and concentration, the final yield was 2.74 mg/L, which was higher than that observed for either 1BKR or 2GKG tags.

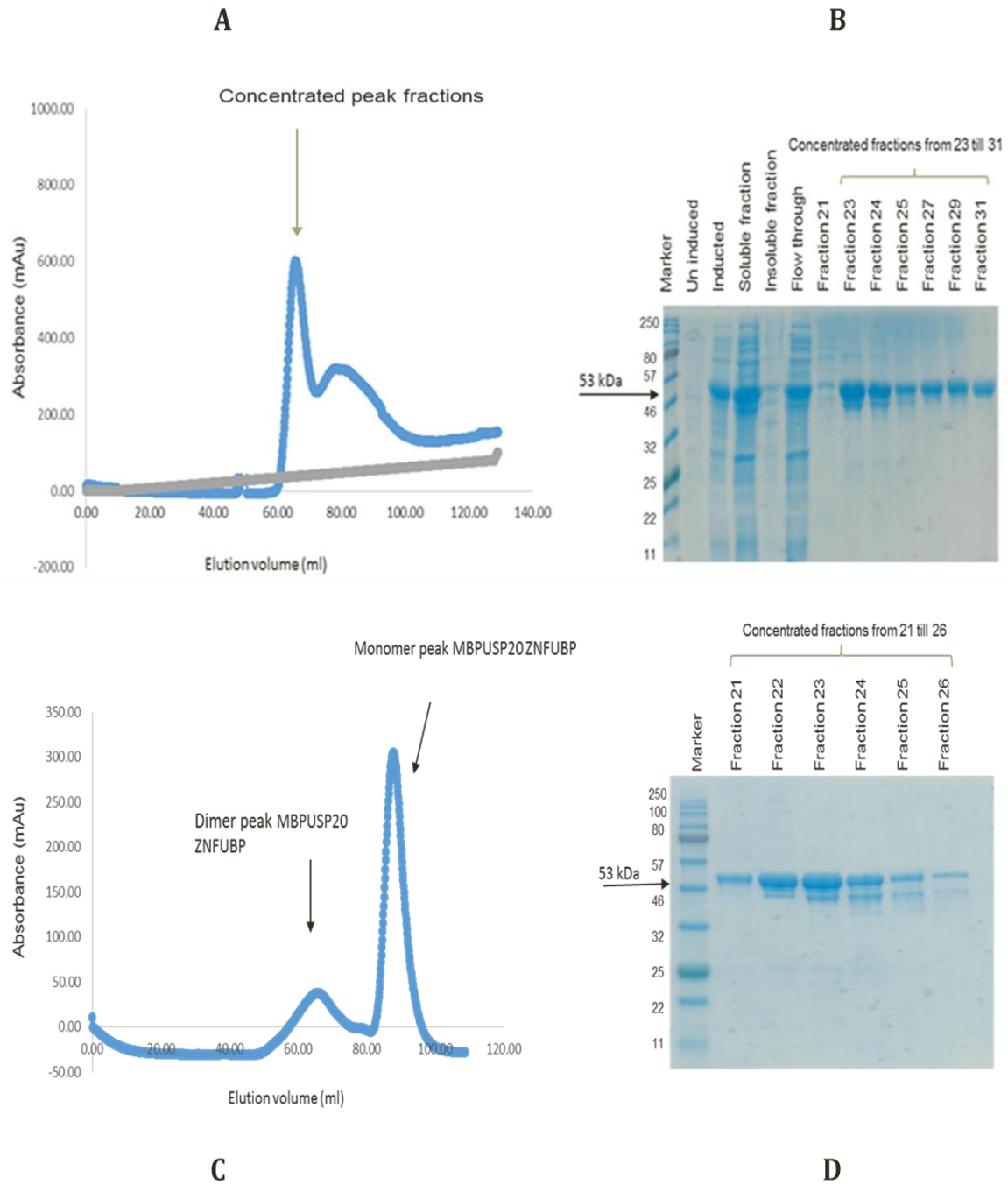


Figure 66. A: Nickel column chromatogram showed a single peak. **B:** SDS-PAGE gel showing the MBP-USP20ZNFUBP protein was expressed. **C:** Gel filtration chromatogram of MBP-USP20ZNFUBP (1-108). **D:** The SDS-PAGE gel of MBPUSP20ZNFUBP showed the single protein was purified.

6.1.4 Crystallisation trials of USP20ZNFUBP(1-108) protein-solubility tag fusions

Crystallization trials were performed to obtain 1BKR-USP20 ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and MBP-USP20ZNFUBP (1-108) crystals. These trials were carried out using concentrations of 1 mg/mL, 2.5 mg/mL and 5 mg/mL of purified 1BKR-USP20ZNFUBP(1-108) and MBP-USP20ZNFUBP(1-108) proteins. On the other hand, for 2GKG-USP20ZNFUBP (1-108) 1 mg/mL and 3.4 mg/mL of purified protein were used. Five different screens were used such as Morpheus, Stura Footprint Combination™ HT-96 code (molecular dimesnion1-43), PACT Premier, PEG Suite II, and JCSG plus. After more than one-week incubation at 20°C, there were no crystals observed for 1BKR-USP20 ZNFUBP (1-108) and MBP-USP20ZNFUBP (1-108) proteins but for 2GKG-USP20ZNFUBP (1-108) protein revealed some objects similar to crystals were obtained on Morpheus screen (Figure 69) but no diffraction was obtained. The crystal screens result after one week are shown in tables 24, 25 and 26.

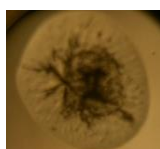


Figure 67. Heavy precipitate in JCSG plus screen from 1BKR-USP20 ZNFUBP protein. The condition was 0.2 M Zinc acetate dihydrate 0.1 M Sodium acetate, 10 % w/v PEG 3000, pH 4.5.

Crystallization trials: 1BKR-USP20 ZNFU(1-108) protein 5 mg/ml			
screens	Clear	Light precipitate	Heavy precipitate
PEG Suite II	32	25	39
JCSG plus	12	36	48
PACT Premier	39	30	27
Morpheus	19	23	54
Stura Footprint	43	38	15

Table 24. Results of crystallisation using different crystal screens for 5 mg/mL 1BKR-USP20 ZNFUBP (1-108) protein after one week.



Figure 68. Crystals from the 2GKG-USP20 ZNFUBP (1-108) protein on Morpheus screen condition: 0.12 M Ethylene glycols 0.1 M Buffer System 1, 50 % v/v Precipitant Mix 1, pH 6.5.

Crystallization trials 2GKG-USP20 ZNFU (1-108) protein 3.4 mg/ml			
screens	Clear	Light precipitate	Heavy precipitate
PEG Suite II	9	39	48
JCSG plus	21	50	25
PACT Premier	18	59	19
Morpheus	18	21	57
Stura Footprint	19	45	32

Table 25. Results of crystallisation using different crystal screens for 3.4 mg/mL 2GKG- USP20 ZNFUBP (1-108) protein after one week.

Crystallization trials: MBP-USP20ZNFU (1-108) protein 7.5 mg/ml			
screens	Clear	Light precipitate	Heavy precipitate
JCSG plus	12	36	48
PACT Premier	5	55	36
Morpheus	13	17	66

Table 26. Results of crystallisation using different crystal screens for 7.5 mg/mL MBP-USP20ZNFUBP (1-108) protein after one week.

6.2 Investigation of the BBS5 (1-270) protein

This section focused on investigations and comparisons of novel protein tags (1BKR and 2GKG tags) with the second model protein BBS5 (1-270) (unknown structure). The comparisons were done as performed with USP20ZNFUBP residues (1-108) protein, which was in terms of expression, solubility, purification and crystallisation.

6.2.1 Cloning of 1BKR-BBS5 (1-270), 2GKG BBS5 (1-270) and MBP- BBS5 (1-270) constructs

Three constructs of BBS5 (1-270) were generated 1BKR-BBS5 (1-270), 2GKG-BBS5(1-270) and MBP-BBS5 (1-270). The first construct was 1BKR-BBS5 (1-270) cloned with a histidine tag at the N terminus using restriction sites *Not*I and *Xho*I in the pRSF-13 vector.

The second construct was 2GKG-BBS5 was cloned using the pRSF-13 vector which inserted a 2GKG tag, a histidine tag at the N terminus.

The MBP- BBS5 (1-270) construct was created by restriction enzyme free cloning protocol as mention previously (material and methods chapter). The 2GKG-BBS5 (1-270) construct was used to engineer MBP-BBS5(1-270) construct by substituting the 2GKG tag with MBP tag. The sequencing results confirmed that MBP-BBS5 (1-270) had been engineered correctly.

6.2.2 Expression, solubility and small scale purification of 1BKR-BBS5 (1-270), 2GKG-BBS5 (1-270) and His-BBS5 (1-270) protein

This experiment investigated the role and influence of designed tags with BBS5 (1-270) protein, in terms of expression, solubility and purification at a small scale. This experiment compared three different constructs 1BKR-BBS5 (1-270), 2GKG-BBS5 (1-270) and His-BBS5 (1-270).

Analysis by SDS-PAGE confirmed all the proteins had been expressed and there were bands in the post-induction fractions (Figure 70 A). All post-induction bands for each protein were of the correct size which were His-BBS5 (1-270) (31 kDa), 2GKG-BBS5(1-270) (45.2 kDa) and 1BKR BBS5 (1-270) (44.2 kDa). The band of 2GKG-BBS5 (1-270) after induction was more pronounced than both bands observed with His-BBS5 (1-270) and 1BKR- BBS5 (1-270), which indicated that 2GKG tag had higher effect on the expression of BBS5 (1-270) protein.

SDS -PAGE analysis revealed the bands of the soluble fractions of 2GKG-BBS5 (1-270) were more pronounced than that observed for 1BKR-BBS5 (1-270) and His-BBS5 (1-270) (Figure 70 B). SDS-PAGE analysis revealed that the presence of 2GKG enhanced the solubility of BBS5 (1-270) protein.

SDS PAGE analysis revealed the purification band of the 2GKG tag was more intense than that observed for the other tags (Figure 70 C). This suggesting the presence of a 2GKG tag improved the purification of BBS5 (1-270) protein when compared with the presence of other tags.

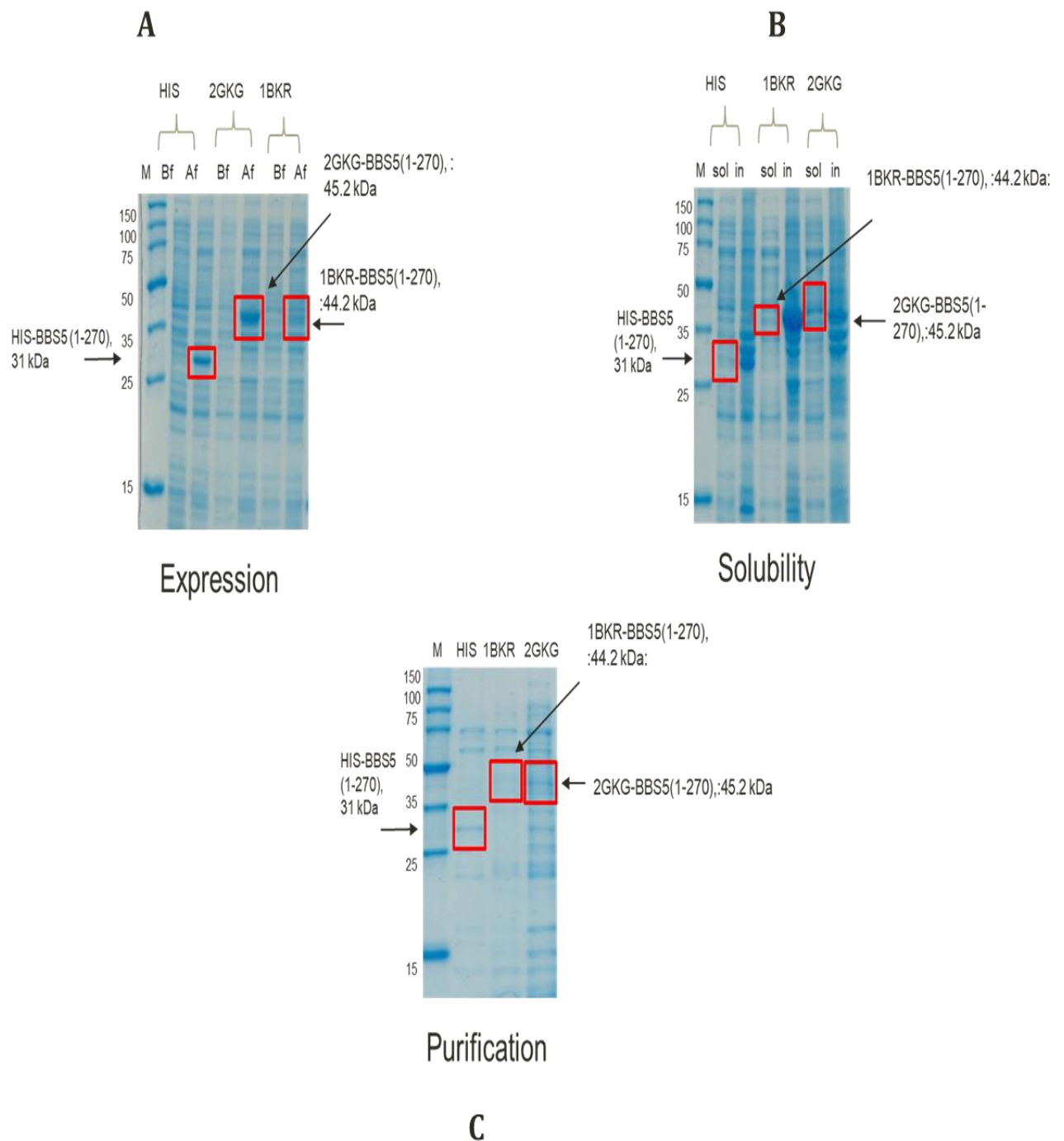


Figure 69. A: The SDS-PAGE gel of expression test of the three proteins His-BBS5 (1-270), 2GKG-BBS5 (1-270) and 1BKR-BBS5 (1-270). B: The SDS-PAGE gel of solubility test of the three proteins His-BBS5 (1-270), 2GKG-BBS5 (1-270) and 1BKR-BBS5 (1-270). C: The SDS-PAGE gel of purification test of the three proteins His-BBS5 (1-270), 2GKG-BBS5 (1-270) and 1BKR-BBS5 (1-270).

6.2.3 Large-scale expression and purification of BBS5 (1-270) protein-solubility tag fusions

6.2.3.1 1BKR-BBS5 (1-270) purification

The purification using Nickel affinity of 1BKR-BBS5(1-270) protein showed a degradation peak (Figure 71 A). All the fractions of the peaks were analysed on SDS-PAGE gel indicating that the protein had been purified with some contamination (Figure 71 B). The 1BKR-BBS5 (1-270) protein bands appeared below 55.6 kDa band and above 42 kDa band on the protein marker (Figure 71 B).

The purification by gel filtration using a S200 Superdex column, the chromatogram showed an elution volume for the protein at 86 mL with a small peak after the void volume (Figure 71 C). This elution volume corresponded to a protein of approximately 44 kDa based on the standard curve. The SDS-PAGE gel showed the final protein yield was too low, as the bands were faint at the correct size for 1BKR-BBS5 (1-270) protein of 44 kDa (Figure 71 D). Some degradation might have occurred which results in the low concentration observed. The final expression yield was 0.1 mg/L.

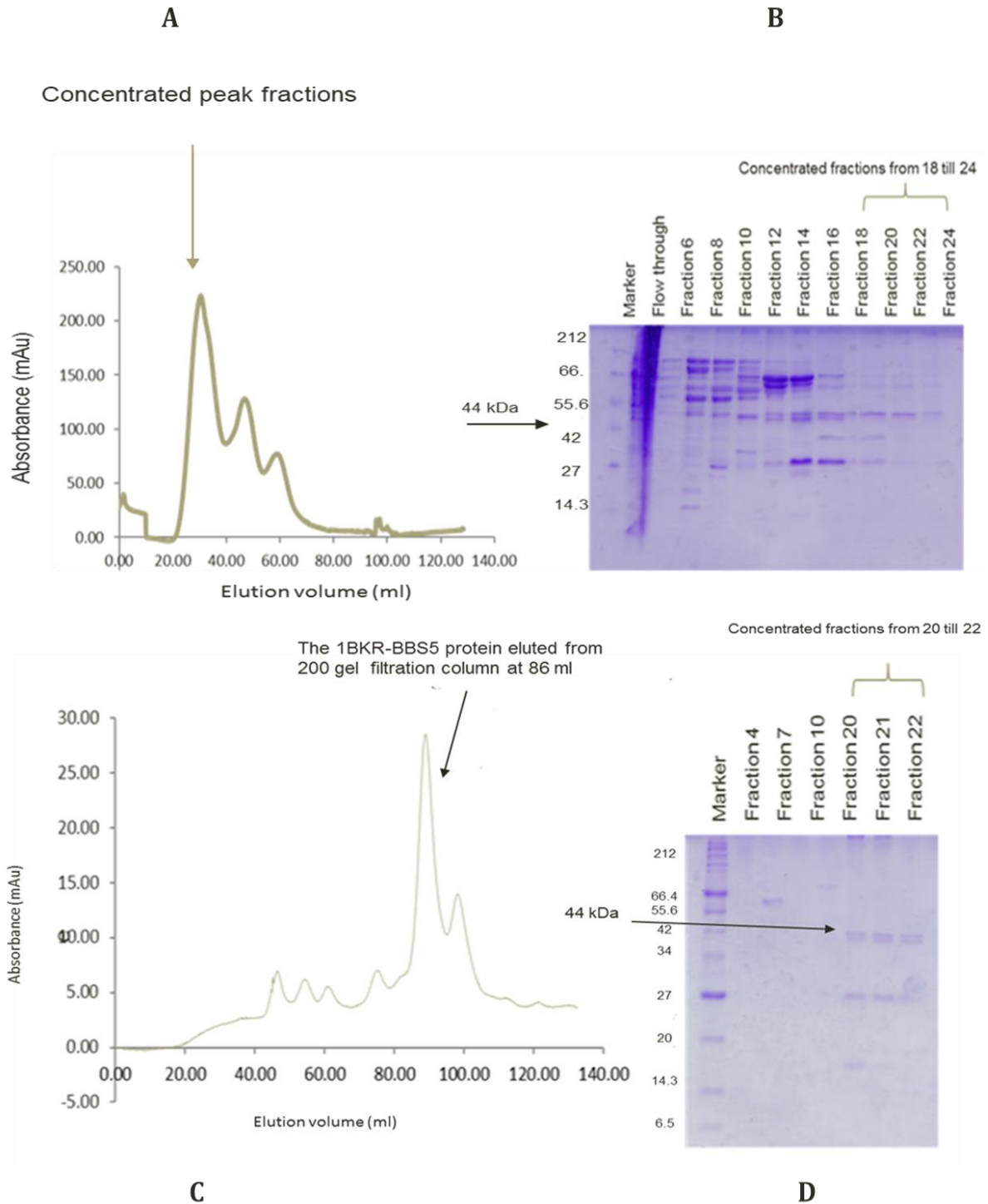


Figure 70.A: Nickel column chromatogram of 1BKR-BBS5 (1-270).B: The SDS-PAGE gel indicates contaminate with 1BKR-BBS5 (1-270) protein..C: The chromatogram of gel filtration showed an elution volume of protein at 86 mL with a small peak after the void volume. D: Th SDS-PAGE gel after gel filtration from 1BKR-BBS5 (1-270).

6.2.3.2 2GKG-BBS5 (1-270) purification

The Nickel affinity purification showed two peaks (Figure 72 A). The both peaks fractions were collected and run on SDS-PAGE gel and analysis revealed that the protein was purified but with some impurities is the second peak (Figure 72 B).

The chromatogram of gel filtration purification showed two peaks (Figure 72 C). The first peak was the void volume peak at 45 mL and the second peak was the correct elution volume for the target protein of 85 mL when compared with the standard curve. This volume refers to about 45 kDa. The SDS-PAGE gel showed single band of the correct size for the target protein under the 55 kDa band of the marker (Figure 72 D). The total protein yield was 0.6 mg/L culture.

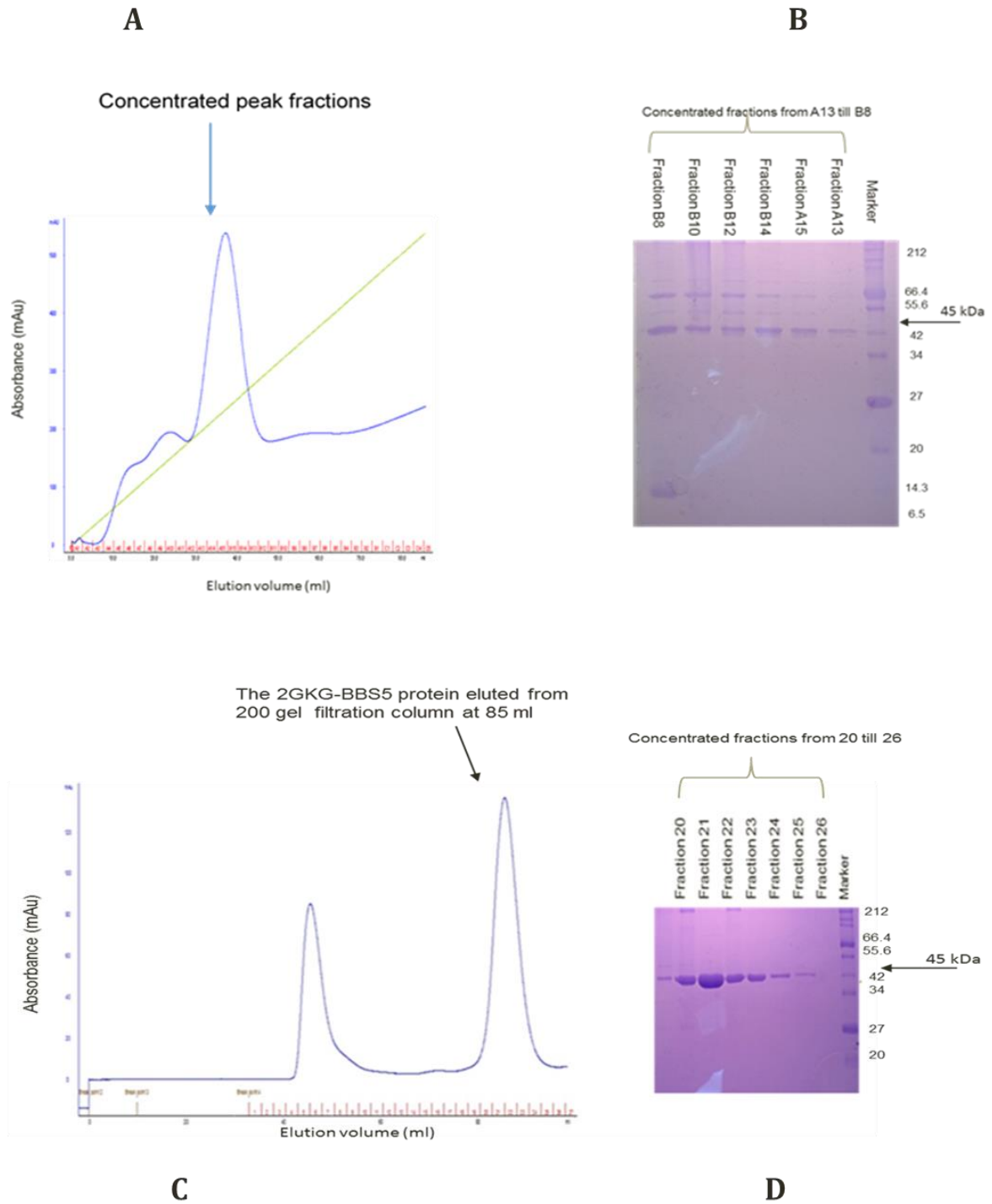


Figure 71A: Nickel column chromatogram of 2GKG-BBS5 (1-270). **B:** The SDS-PAGE gel after Nickel column purification of 2GKG-BBS5 (1-270) protein. **C:** The gel filtration chromatogram of 2GKG-BBS5 (1-270) protein. **D:** The SDS-PAGE gel after gel filtration purification of 2GKG-BBS5 (1-270) protein.

6.2.3.3 MBP-BBS5 (1-270) purification

The nickel column purification showed two peaks (Figure 73 A). The protein target present in the first peak as the SDS PAGE gel confirmed the protein target with a protein band of approximately 71 kDa (Figure 73 B). The protein MBP-BBS5 (1-270) has a predicted size of 71 kDa.

The chromatogram of gel filtration purification showed four peaks (Figure 73 C), analysis showed that the protein of interest was present in the fourth peak, this peak eluted at 77 mL. This elution volume refers to 71 kDa based on the standard curve showed that MBP-BBS5 (1-270) in the monomer formation. This size was almost equal to the actual size of the protein MBP-BBS5 (1-270). SDS PAGE gel showed a pure protein product of the correct molecular weight (Figure 73 D). The total final yield was 0.42 mg/L culture.

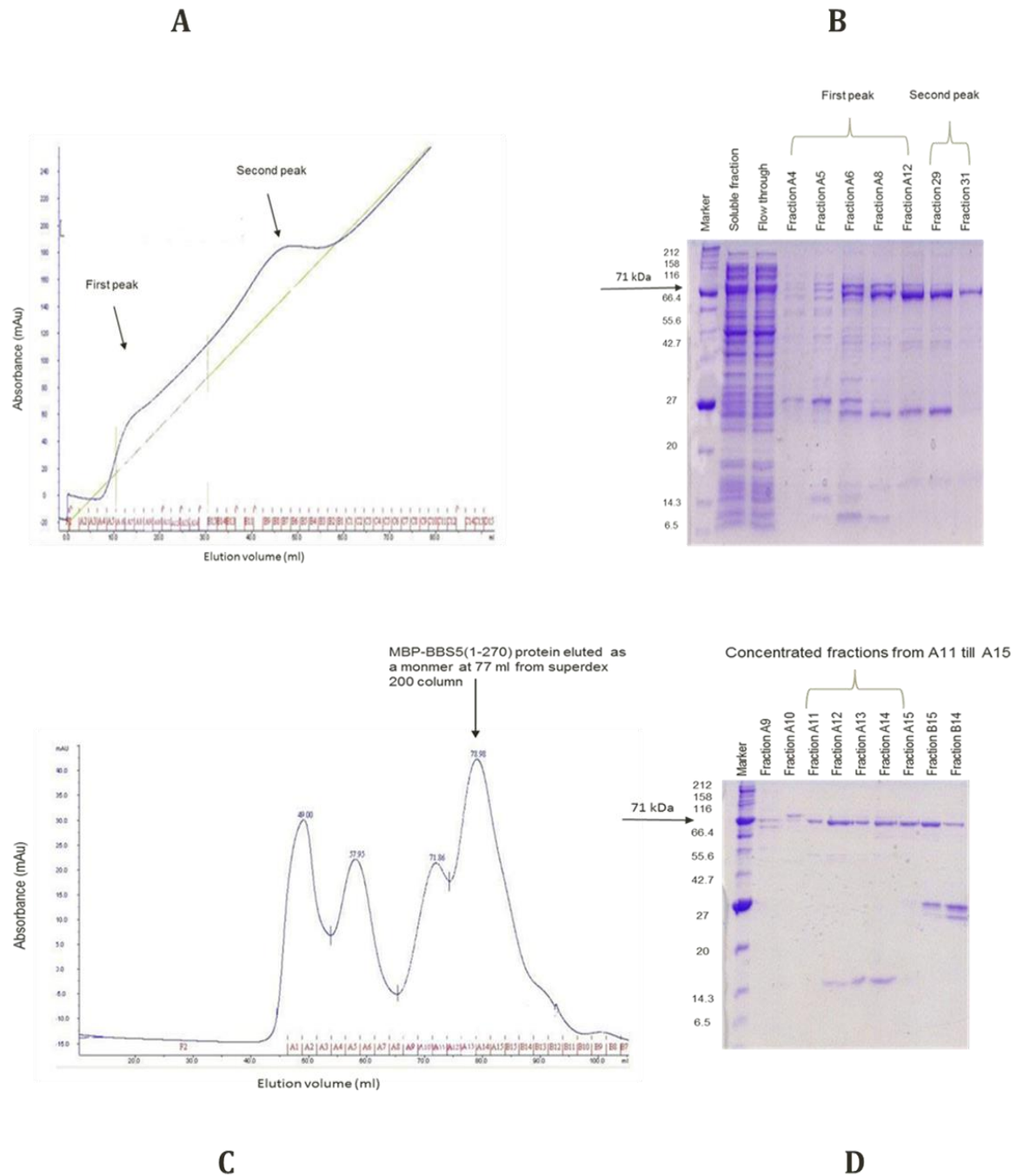


Figure 72. A: The affinity chromatography chromatogram of MBP-BBS5 (1-270). B: SDS-PAGE after Nickel column purification of MBP-BBS5 (1-270). C: The chromatogram of gel filtration purification of MBP-BBS5 (1-270). D: SDS-PAGE gel result showing bands that correspondent to MBP-BBS5 (1-270) (71 kDa).

6.2.4 Crystallisation trial BBS5 (1-270) protein- solubility tag fusions

The Crystallisation trials of protein 1BKR-BBS5 (1-270), 2GKG-BBS5 (1-270) and MBP-BBS5 (1-270) were performed using different screens such as Morpheus, PEG Suite II, and JCSG plus with 1 mg/mL protein concentration. After incubation for 3 months at 20°C no crystals were obtained. The crystal screens result after one week are in tables 27 and 28.

Crystallization trials: 1BKR-BBS5 (1-270) protein 1 mg/mL			
screens	Clear	Light precipitate	Heavy precipitate
JCSG plus	51	23	22
PEG Suite II	39	41	16
Morpheus	45	23	28

Table 27. Results of crystallisation using different crystal screens for 1 mg/mL 1BKR-BBS5 (1-270) protein after one week.

Crystallization trials :2GKG -BBS5 (1-270) protein 1 mg/mL			
screens	Clear	Light precipitate	Heavy precipitate
JCSG plus	38	42	16
PACT premiere	54	23	19
Morpheus	39	22	35

Table 28. Results of crystallisation using different crystal screens for 1 mg/mL 2GKG -BBS5 (1-270) protein after one week.

6.3 Discussion of the results of USP20ZNFUBP (1-108) and BBS5 (1-270) fusions with the tags

USP20ZNFUBP (1-108) and BBS5 (1-270) are difficult proteins to express and purify and their structures are not known. This chapter showed the results of the novel tags 1BKR, and 2GKG when fused with USP20ZNFUBP (1-108) and BBS5 (1-270) proteins. The performance of these novel tags to express and purify these proteins in small and large scales were compared with the commonly used tags such as Histidine and MBP.

The results from small-scale SDS-PAGE analysis indicated that the 2GKG and 1BKR tags improved the expression, solubility and purification of USP20ZNFUBP when compared with the presence of a His tag. However, results for the BBS5 (1-270) protein indicated that although the 1BKR tag was comparable to the His tag results, the presence of 2GKG significantly improved protein expression. These results suggested that the 2GKG tag had a higher effect on the yield of both USP20ZNFUBP (1-108) and BBS5 (1-270) proteins than the presence of either a 1BKR tag or Histidine tag.

At the large scale, results showed that all the proteins expressed and purified well but the USP20ZNFUBP (1-108) protein with 2GKG tag was not stable and aggregated, while the 1BKR tag was stable with USP20ZNFUBP (1-108) resulting in a higher yield. The MBP tag results showed it has the highest effect on USP20ZNFUBP (1-108) protein when compared to 1BKR and 2GKG tags as it obtained the highest soluble protein yield (Table 29 and Figure 74).

For the BBS5 (1-270) protein at the large scale expression and purification trial, the 2GKG tag had the highest effect on BBS5 (1-270) protein when compared to MBP and 1BKR tags as the 2GKG had the highest yield (Table 30 and Figure 75). This may be due to its small molecular weight when compared with MBP tag and more favorable steric and electrostatic effects. This effect has been observed when comparing a SUMO tag with a MBP tag fused to Growth differentiation factor 8(GDF8) protein, which suggested that the presence of a small tag is favorable in some cases (262) and that was one of the criteria when selecting these novel tags as they all possess a small molecular weight and are less likely to lead to steric clashes. The results from the small and large-scales trials indicated that the novel tags 1BKR and 2GKG could efficiently be used as solubility tags for the purpose of expressing and purifying difficult proteins such as USP20ZNFUBP (1-108) and BBS5 (1-270). The results were comparable to the Histidine and MBP tags which are most efficient fusion tag in many studies (263).

With regards to crystallization, all tags failed to form crystals. This is may be due to the low concentration of protein in the drops which was not enough to reach the super-saturation state and form crystals, as crystal formation require high protein concentration, although the solubility of these proteins appeared low (264).

The second possibility may be due to the incubation temperature (20°C) which can be varied for proteins to form crystals, which is one of the factors affecting protein crystallization (265).

The third possibility was that USP20ZNFUBP (1-108) and BBS5 (1-270) proteins have some disordered regions which affect crystal formation. These proteins sequences were investigated by Swiss model and xtalpred servers detecting disordered

regions in both proteins of unknown structure and the possibility of protein crystallization (266) (267). Both servers revealed that USP20ZNFUBP (1-108) and BBS5 (1-270) proteins contain some disordered regions (data not represented) which might hinder crystal formation.

The construct	Yield (mg/L)
1BKR- USP20 ZNFU (1-108)	1.25
2GKG-USP20 ZNFU (1-108)	0.86
MBP- USP20 ZNFU (1-108)	2.7

Table 29.Comparison of yields of 1BKR-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP (1-108) and MBP-USP20ZNFUBP (1-108) constructs after gel filtration purification.

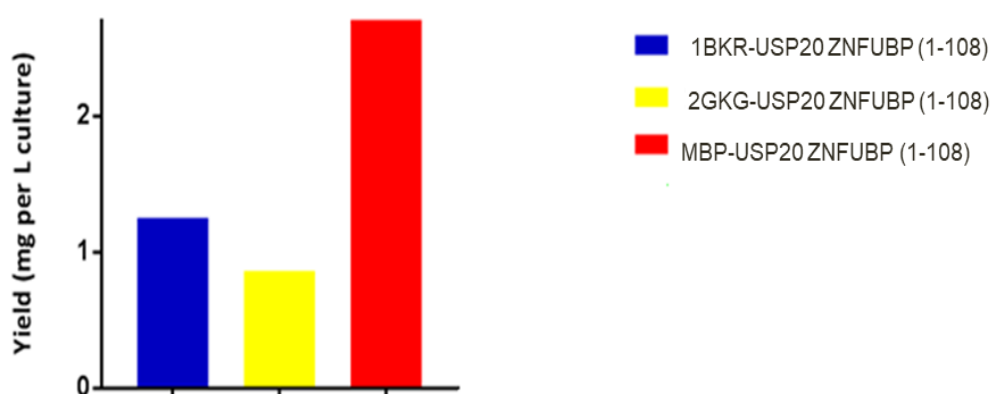


Figure 73.Comparison of yields of 1BKR-USP20ZNFUBP (1-108), 2GKG-USP20ZNFUBP(1-108) and MBP-USP20ZNFUBP (1-108) constructs after gel filtration purification (no error bar because it just only one experiment).

The construct	Yield (mg /L)
1BKR-BBS5 (1-270)	0.06
2GKG BBS5(1-270)	0.6
MBP-BBS5 (1-270)	0.42

Table 30.Comparison of yield of 1BKR- BBS5 (1-270) 2GKG- BBS5 (1-270), MBP- BBS5 (1-270) constructs after gel filtration purification.

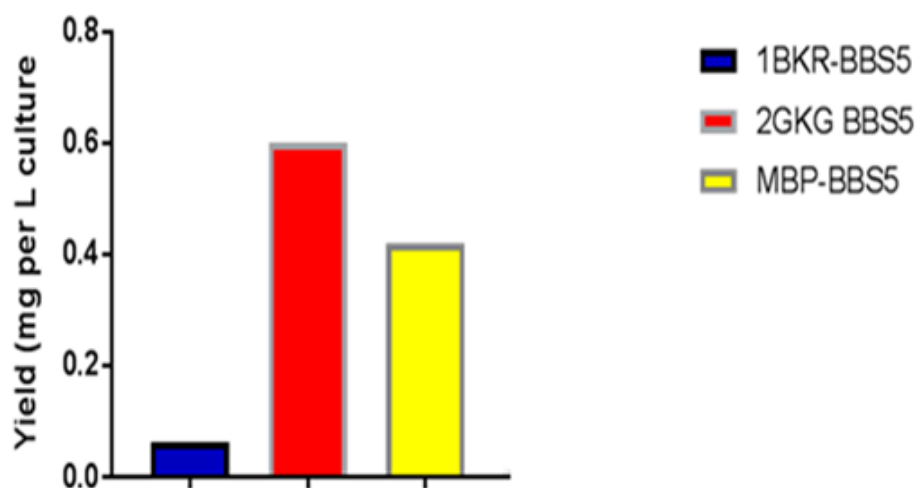


Figure 74.Graph showed the comparison of yields of 1BKR- BBS5 (1-270), 2GKG- BBS5 (1-270), MBP- BBS5 (1-270) constructs after gel filtration purification (no error bar because it just only one experiment).

7. Chapter 7: Results and Discussion of Ubiquitin protein applications as affinity, solubility tag or protein scaffold

Existing fusion tags cannot be used for all proteins leading to the need for the identification of new tags. As a result, the ubiquitin protein was investigated as it has been shown to have high stability and solubility. These properties suggest that this protein could be used as fusion tags according to many studies have shown that protein stability and solubility are important criteria for fusion tags to improve expression and solubility of the fusion proteins (268) (56). Previous study had shown that Small Ubiquitin-like Modifier (SUMO) tag has high stability and solubility which permit it to be used as fusion tag, fusing with different proteins and peptides had an equivalent performance to many commonly used fusion tags such as maltose-binding protein (MBP) and glutathione S-transferase (262). Ubiquitin proteins have also been shown to have the ability to be used as fusion tag as it has the ability to improve the efficient translation of the fusion partner (58). Therefore, ubiquitin protein was further investigated as a fusion tag in modified form.

In addition, some studies showed that ubiquitin due to its high stability and solubility can be engineered for high affinity target binding and could be used as a protein scaffold adaptable to specific tasks as therapeutic purposes (269). In this study, I engineered ubiquitin to be used as an affinity and solubility tag and also to be used as a protein scaffold for displaying constrained peptides for high affinity target binding

7.1 Ubiquitin protein as affinity and solubility tag

A histidine tag was inserted in the loop between the strands 3 and 4 in the ubiquitin protein in order to design this protein as an affinity tag (Figure 76). This approach has been successful with a Green Fluorescent Protein(GFP) tag which purified the fusion protein with Histidine tag in the loop of the GFP tag, which suggests the Histidine tag in the loop in some tags can perform better than located either in the C or N terminus, as it will not be so easily cleaved and will be less flexible (270).

This section will show the cloning of a UbS3S4His tag and then its suitability as affinity tag when fused to two difficult to express proteins USP20 ZNFUBP (1-108) and BBS5 (1-270) for which I had tested the 2GKG and 1BKR tags before.

7.1.1 Cloning of a UbS3S4His tag construct

In order to design the UbS3S4His construct, a *Bam*HI restriction site was engineered into ubiquitin loop between strands 3 and 4 as a new restriction site in addition to the original *Bam*HI site in the loop of the ubiquitin protein between strands 3 and 4. The aim of design new *Bam*HI in the loop to insert Histidine tag is to use annealing oligonucleotides primers cloning protocol which is easier for peptides cloning. The histidine tag was inserted between strands 3 and 4 (Figure 76) as the sequencing result confirmed that the UbS3S4His construct was engineered correctly.

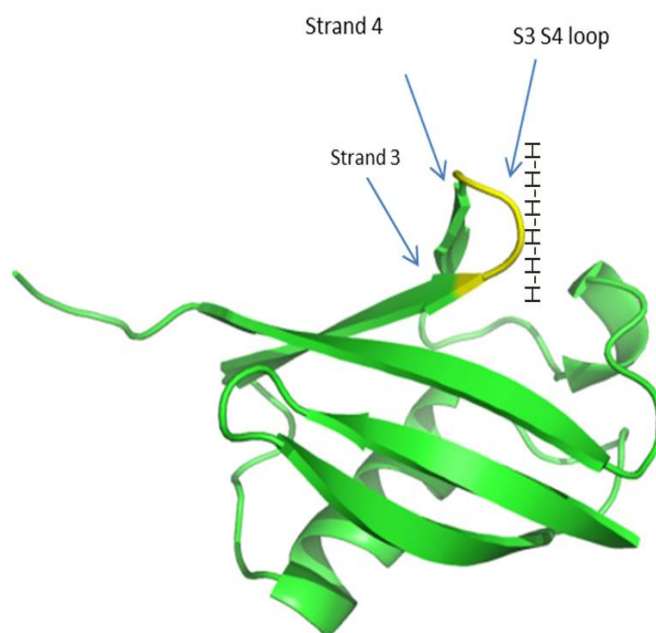


Figure 75 . Cartoon diagram of the ubiquitin protein structure (PDB code 1UBI) and the yellow colour represents the loop between strands 3 and 4 where the histidine tag was inserted.

7.1.2 UbS3S4His expression and purification

The purification method of UbS3S4His involved affinity chromatography as the first purification step and gel filtration as second purification step for assessment of the protein solubility and stability. The chromatogram of nickel chromatography purification showed a double peak (Figure 77 A) with the protein target in the second peak as indicated by analysis of a SDS-PAGE gel (Figure 77 B). Analysis of flow through fraction and the peak fractions by SDS-PAGE gel indicated the UbS3S4His bound with nickel column, as there was no band in the flow through at 8.5 kDa (Figure 77 B). These results suggested that UbS3S4His protein could be used as an affinity tag as the in the loop inserted histidines bound to the column.

The final purification step of UbS3S4His was gel filtration, the protein eluted at 95 mL (Figure 77 C). This volume was equal to 8.5 kDa on the standard curve of the

superdex 75 column. These results showed that the protein is stable and soluble as the protein had not eluted in the void volume. Peak fractions analysis by SDS-PAGE gel showed a pure protein of approximately 8.5 kDa (Figure 77 D). The final yield obtained was 15 mg/L, which was used for crystallization trials.

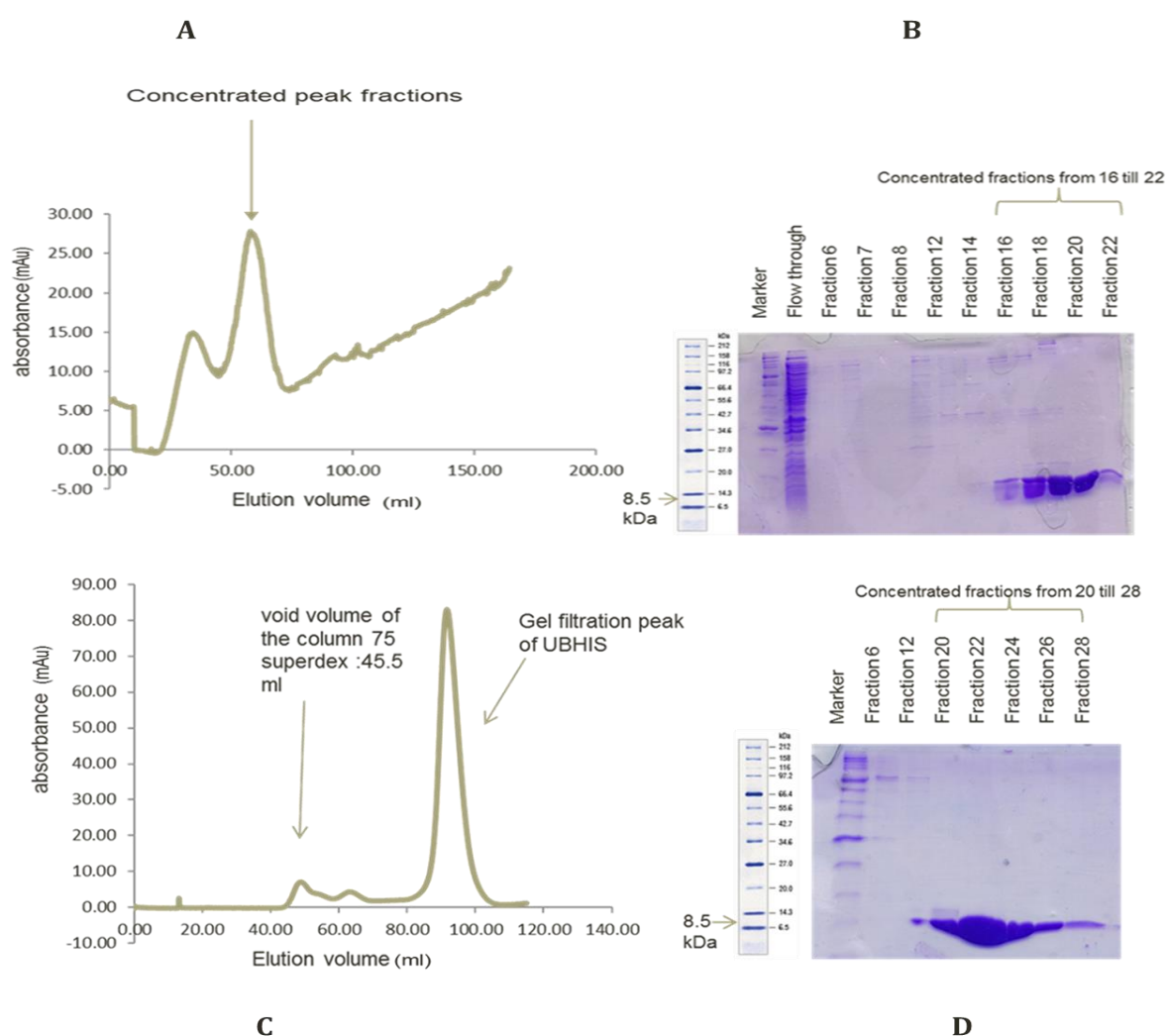


Figure 76. A: The Nickel column purification chromatogram of UbS3S4His. **B:** SDS-PAGE gel result showed bands that corresponded to UbS3S4His at about 8.5 kDa. **C:** The gel filtration chromatogram of UbS3S4His. **D:** SDS-PAGE gel result showed bands that corresponded to UbS3S4His at about 8.5 kDa.

7.1.3 Assessing UbS3S4His protein thermal stability by Thermoflour assay

A stability test (Thermoflour assay) was carried out, which investigated if the insertion of the histidine tag in the loop between strands 3 and 4 affected the thermal stability of the ubiquitin protein. Results revealed that UbS3S4His and ubiquitin proteins had a high melting temperature of at least 92°C. Due to the high stability of both proteins the data were not easily interpretable as there were no peaks observed as usual with Thermoflour results. Peaks are the consequence of protein denaturation but the highest temperature used in the qPCR machine was not high enough to denature both proteins as a result no peaks were observed (Figure 78). This results showed that the insertion of the histidine tag had not affected ubiquitin stability (271).

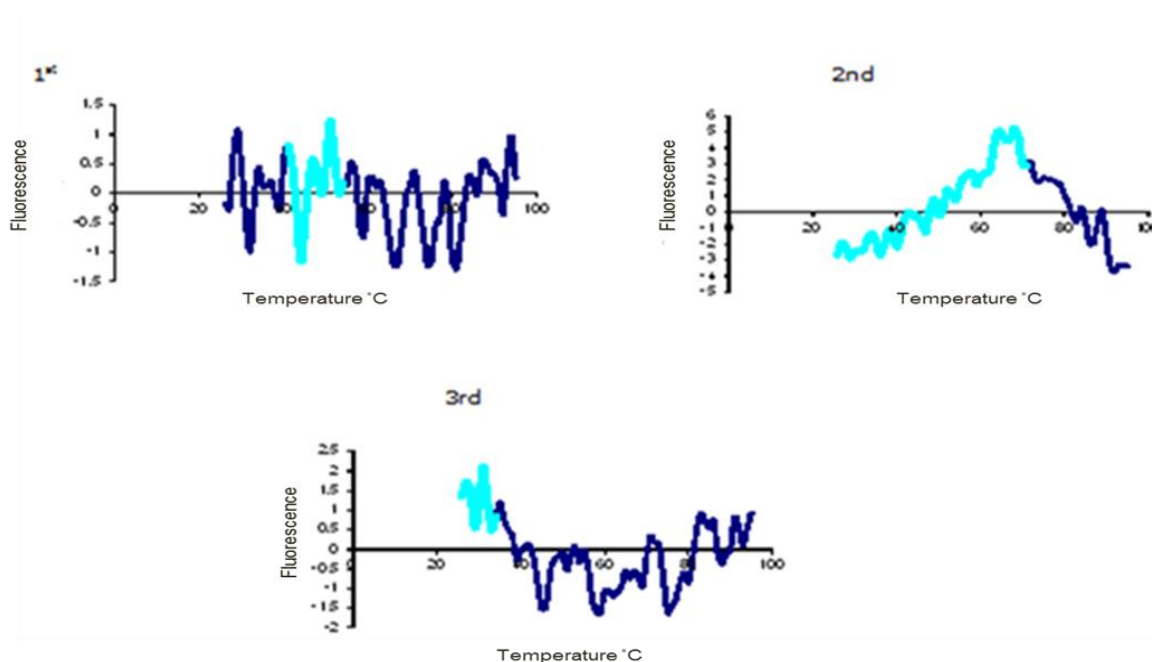


Figure 77. Sample of raw data from the Thermoflour assay. The graph of the fluorescence against temperature of UbS3S4His

7.1.4 Crystallization trials of UbS3S4His

Crystallization trials were performed on the UbS3S4His purified proteins with 2.5-10 mg/mL in different screens such as Stura Footprint Combination™ HT-96 code (molecular dimesnion1-43), PACT premier™, HT-96 code (molecular dimesnion1-36), PGA, HT-96 code (molecular dimesnion1-51 and MIDAS™ HT-96 code. No crystals were obtained, which showed that UbS3S4His hasn't the ability to form crystals easily and this may be due to the His tag in the loop affecting the protein conformation which hinders the protein to crystallize (252).

The effect of fusing UbS3S4His protein with two difficult to express proteins USP20ZNFUBP (1-108) and BBS5 (1-270) was explored for an assessment of expression, purification and crystallization.

7.2 Investigation of UbS3S4His-USP20ZNFUBP (1-108) and UbS3S4His -BBS5 (1-270) fusion proteins

7.2.1 Cloning of UbS3S4His-USP20ZNFUBP (1-108) and UbS3S4His-BBS5 (1-270) constructs

The constructs of UbS3S4His-USP20ZNFUBP (1-108) and UbS3S4His-BBS5 (1-270) were created by insertion the USP20ZNFUBP (1-108) and BBS5 (1-270) genes at the multiple cloning sites in the *NheI* and *HindIII* restriction sites in the UbS3S4His-pRSF13 vector. Sequencing results showed that the new two constructs had been created correctly.

7.2.2 Large-scale expression and purification of UbS3S4His-USP20ZNFUBP (1-108)

UbS3S4His-USP20ZNFUBP (1-108) was purified using Nickel column and gel filtration methods. SDS PAGE gel analysis after purification using nickel affinity, revealed that the UbS3S4His-USP20ZNFUBP (1-108) protein had bound to the nickel column, as there no band in the flow through at the same molecular weight of the protein 21.5 kDa. However, only three fractions contained the target protein which indicated that there was a low yield of expressed and purified protein. SDS-PAGE analysis also revealed some degradation which suggested that the target protein was not stable (Figure 79 A).

The second purification step was a gel filtration, which showed the protein target eluted at 73 mL (Figure 79 B). When compared with a standard curve, the elution volume of UbS3S4His -USP20ZNFUBP (1-108) corresponded to a protein of 21.5 kDa

and the protein on SDS PAGE gel had a molecular weight of approximately 21.5 kDa (Figure 79 C). SDS PAGE gel indicated that the protein had a low yield with some impurities in the sample which was not fit for the crystallization trials and so crystallisation trials were not proceeded. The yield of this purification was 0.005 mg/L culture.

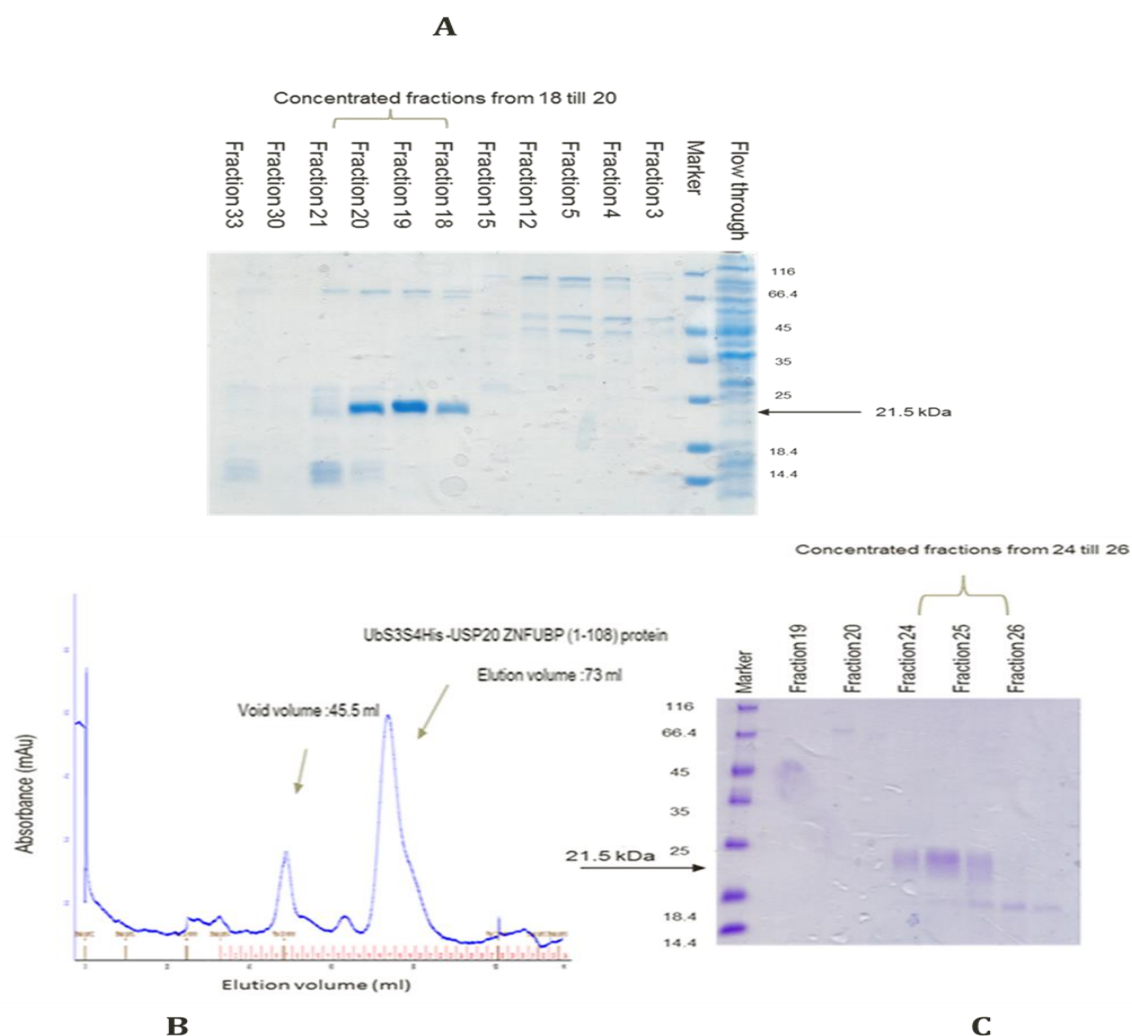


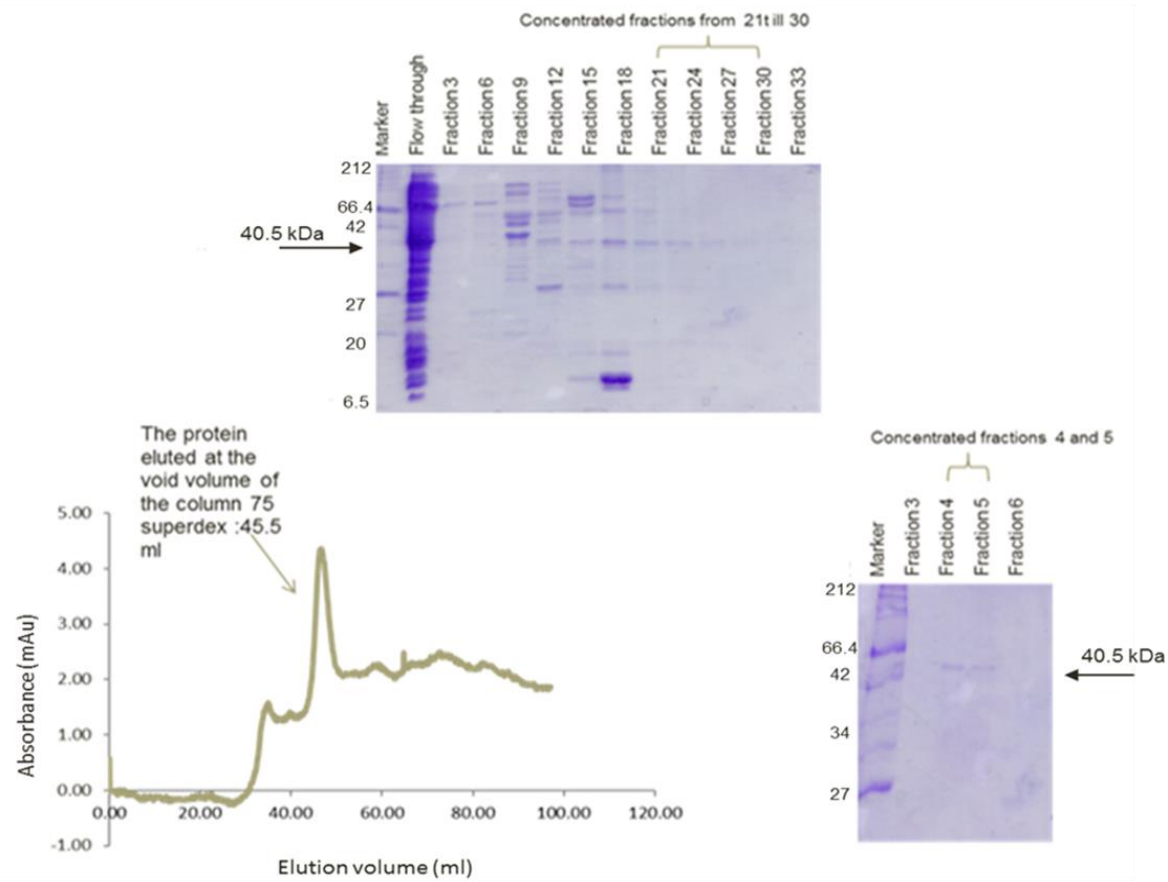
Figure 78.A: SDS-PAGE after the affinity chromatography. **B:** The chromatogram of the gel filtration. **C:** SDS-PAGE gel shows the protein is some degradation product.

7.2.3 Large-scale expression and purification of UbS3S4His-BBS5 (1-270)

The purification was performed by affinity and gel filtration purifications steps. The SDS-PAGE gel after affinity chromatography showed a prominent band in the flow through which indicated the UbS3S4His -BBS5 (1-270) protein had not bound tightly to the nickel column (Figure 80 A). The SDS-PAGE gel confirmed that the protein had been purified at a very low yield as only faint bands of the correct size of the target protein were detected (Figure 80 A). The pure fractions were concentrated and purified by gel filtration for the assesment of the protein stabilty.

The gel filtration purification showed that the protein eluted in the void volume peak with the elution volume of 45.5 mL (Figure 80 B) which indicted the protein had aggregated and was not stable. The fractions of the peak were run on an SDS-PAGE gel which showed that the product had a very low protein concentration (Figure 80 C). No crystallisation trials were performed on UbS3S4His-BBS5 (1-270) due to the protein aggregation. A protein preparation should be pure, soluble and in high concentration to be valid for crystallisation trials. The total pure protein yield after concentration was very low (0.0125 mg/L culture).

A



B

C

Figure 79. A: SDS-PAGE gel after nickel from UbS3S4His -BBS5 (1-270). B: The chromatogram of gel filtration purification of UbS3S4His -BBS5. C: The SDS-PAGE gel also confirmed that the product contains very low protein yield.

7.3 The use of the ubiquitin protein as a protein scaffold

In this section, the ubiquitin protein was examined for the possibility to be used as a protein scaffold for high affinity target binding. Since, the insertion of the Histidine tag into the strands 3 and 4 loop of ubiquitin was successful with no effect on the stability and solubility of the ubiquitin protein, ubiquitin was investigated to be used as a more general scaffold to display constrained peptides in this loop region. Previous research in the Dreveny lab has shown that constrained peptides had been identified from a constrained phage display library which interacted with the USP11 N-terminal domains 1-244 (unpublished results). Therefore, the ubiquitin protein was used as a protein scaffold to investigate the binding of a newly identified peptide with the USP11 residues N-terminal domains (1-244).

I cloned the UbpeptideS3S4 construct and then expressed and purified it as well as USP11 N-terminal domain (1-244) which had been cloned previously.

7.3.1 Cloning of UbpeptideS3S4 and USP11 N-terminal domain (1-244)

The UbpeptideS3S4 construct was engineered by inserting the peptide with the sequence **LASLLPRRS** into the loop between the strands 3 and 4 of the ubiquitin protein by annealing oligonucleotides primers cloning method using *BamHI* restriction sites. For the USP11 N-terminal domain (1-244), an existing construct was used. The gene of USP11 N-terminal domain (1-244) had been inserted in the multiple cloning site with an affinity Histidine tag at the C terminus in pET26b. Inserting a His tag at the C terminus in some cases can produce a higher protein yield than when present at

the N terminus (272). The sequencing result confirmed that the UbpeptideS3S4 was engineered correctly (Figure 81).

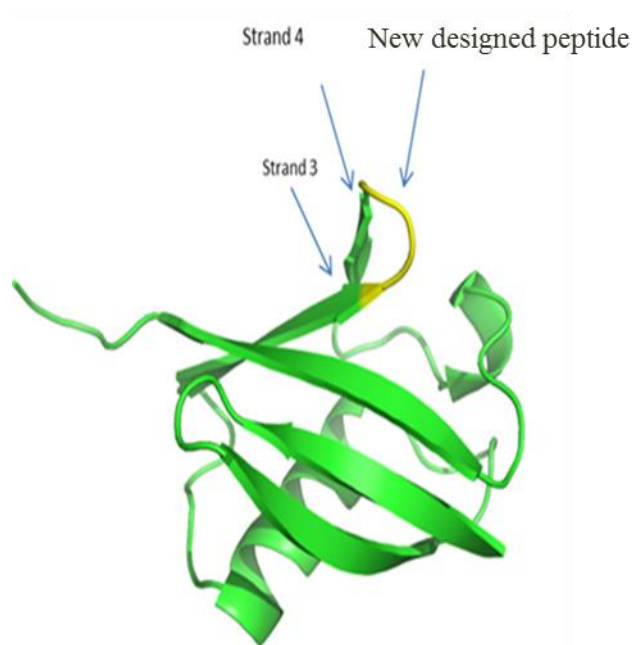


Figure 80. Cartoon diagram of the ubiquitin protein structure (PDB code 1UBI) and the yellow colour represents the loop between strands 3 and 4 where the new peptide was inserted.

7.3.2 Expression and Purification of Ubiquitin peptide S3S4

The aim of this experiment was to obtain a high yield of Ubiquitin peptide S3S4 protein which could then be used for interaction studies. The results showed that the protein was expressed as shown in SDS-PAGE gel as in Figure 82 A. The ubiquitin protein with inserted peptide between strands 3 and 4 was purified by decreasing the pH of *E. coli* lysate and precipitating most of the proteins as ubiquitin protein is very stable and can refold when denatured. The rest of the lysate was loaded onto a column of cation exchange and size exclusion chromatography was used as a second purification step. The chromatogram of the size exclusion chromatography showed the protein eluted at the expected elution volume (Figure 82 B). The peak fractions were

analysed by SDS-PAGE gel, which indicated very pure protein was obtained (Figure 82 C). The final yield was 0.425 mg/L culture.

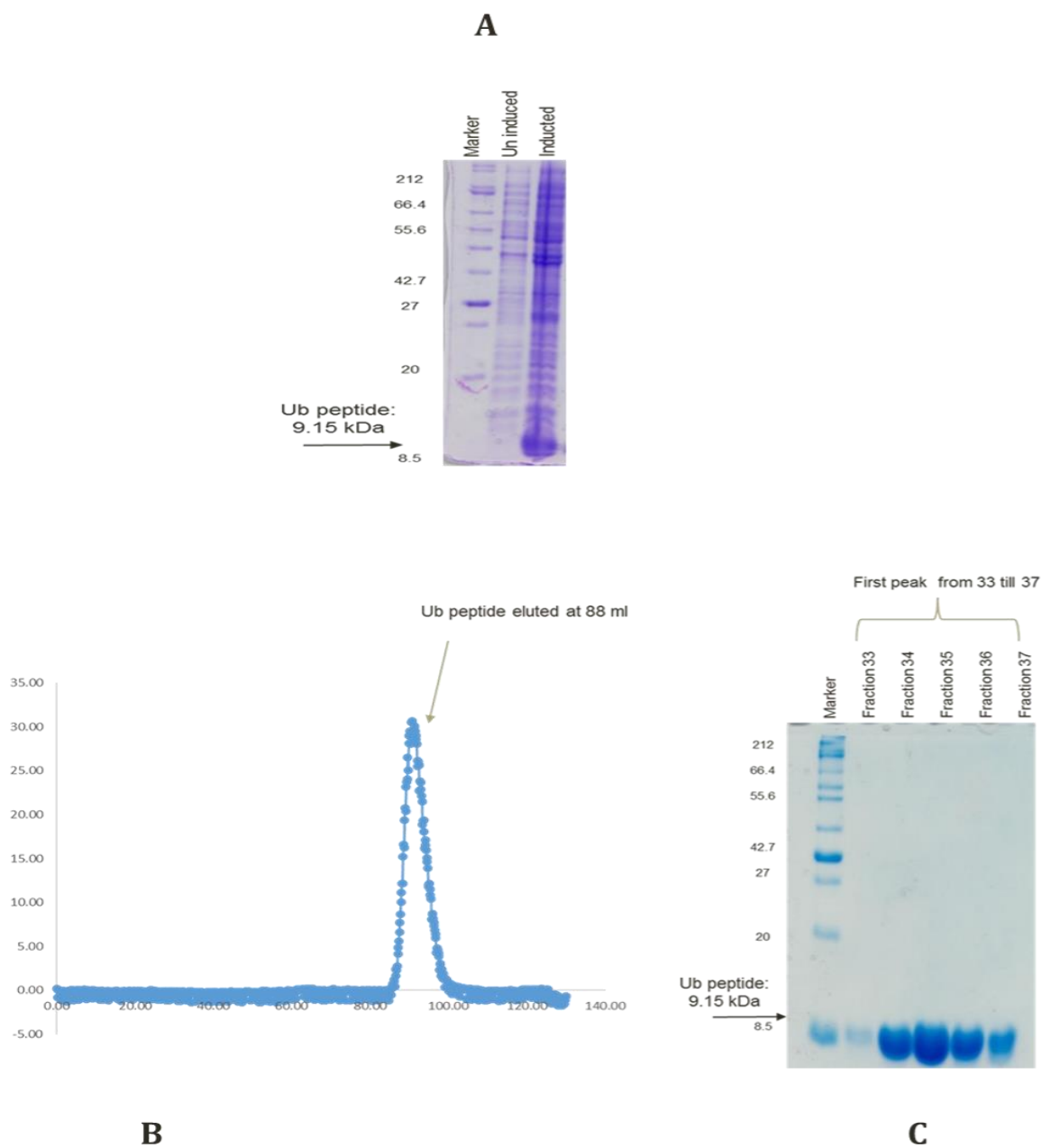


Figure 81. A: SDS-PAGE gel of the expression test of UbpepS3S4. B: The chromatogram of the size exclusion chromatography C: SDS-PAGE gel after gel filtration from UbpepS3S4 of the gel filtration.

7.3.3 Expression and purification of USP11 domain residues (1-244)

The USP11 (1-244) protein was expressed and purified by Nickel affinity and gel filtration chromatography. The result of the Nickel affinity purification indicated that the protein had been well expressed. The purification of USP11(1-244) showed a broad single peak (Figure 83 A). The peak fractions were analysed on SDS-PAGE gel showed the target protein USP11 (1-244) with a molecular weight of 28.6 kDa, as an intense band above 27 kDa of the marker band (Figure 83 B).

The gel filtration purification of USP11 (1-244) eluted as a single peak at an elution volume of 67 mL (Figure 83 C). This elution volume equated to the volume of a standard curve at the prediction of molecular weight of 28.6 kDa. The SDS-PAGE gel confirmed the purity of protein (Figure 83 D) and yield was 1.58 mg/L culture.

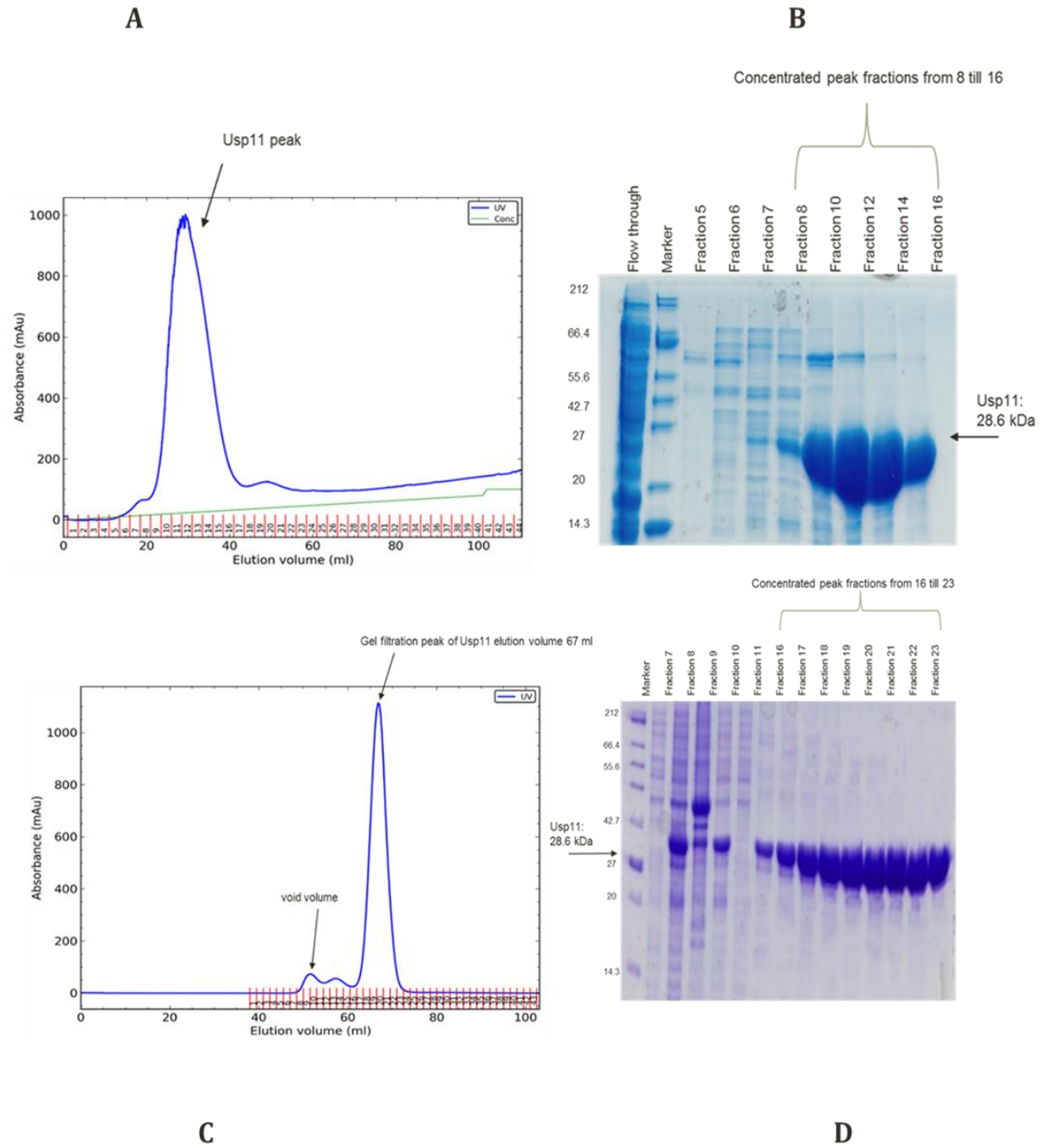


Figure 82. A: The Nickel column purification chromatogram of USP11 residues (1-244). B: SDS-PAGE gel after Nickel column purification C: Gel filtration chromatogram of USP11 residues (1-244). D: SDS-PAGE gel after gel filtration purification of USP11 residues (1-244).

7.3.4 Interaction studies of USP11 (1-244) with UbpepS3S4 by isothermal titration calorimetry

Isothermal titration calorimetry was used to evaluate the binding between USP11 residues 1-244 and UbpeptideS3S4 proteins. Both the USP11 residues 1-244 and ubiquitin peptide were run on a gel filtration Superdex 75 column, as showed before using the same buffer to prevent buffer mismatch (170). The ITC results showed a binding curve (Figure 84) with an estimated equilibrium dissociation constant (K_d) value of about 12.2 μ M and the binding ratio of $N=0.6$ which showed that in the calorimetric cell 60% of USP11 protein was active (binding competent) or may reflect inaccuracies in the protein concentration measurements. As the ubiquitin construct does not contain any tryptophans, concentrations can be inaccurate. The equilibrium dissociation constant (K_d) value was in micromolar not in nanomolar which indicating the interaction is not so tight but for protein-peptide interactions this is not unusual (273) (274). These results suggest that the Ubiquitin protein can be used as scaffold but it needs further investigation by using different proteins as controls, allowing a comparison with USP11 (1-244) protein result.

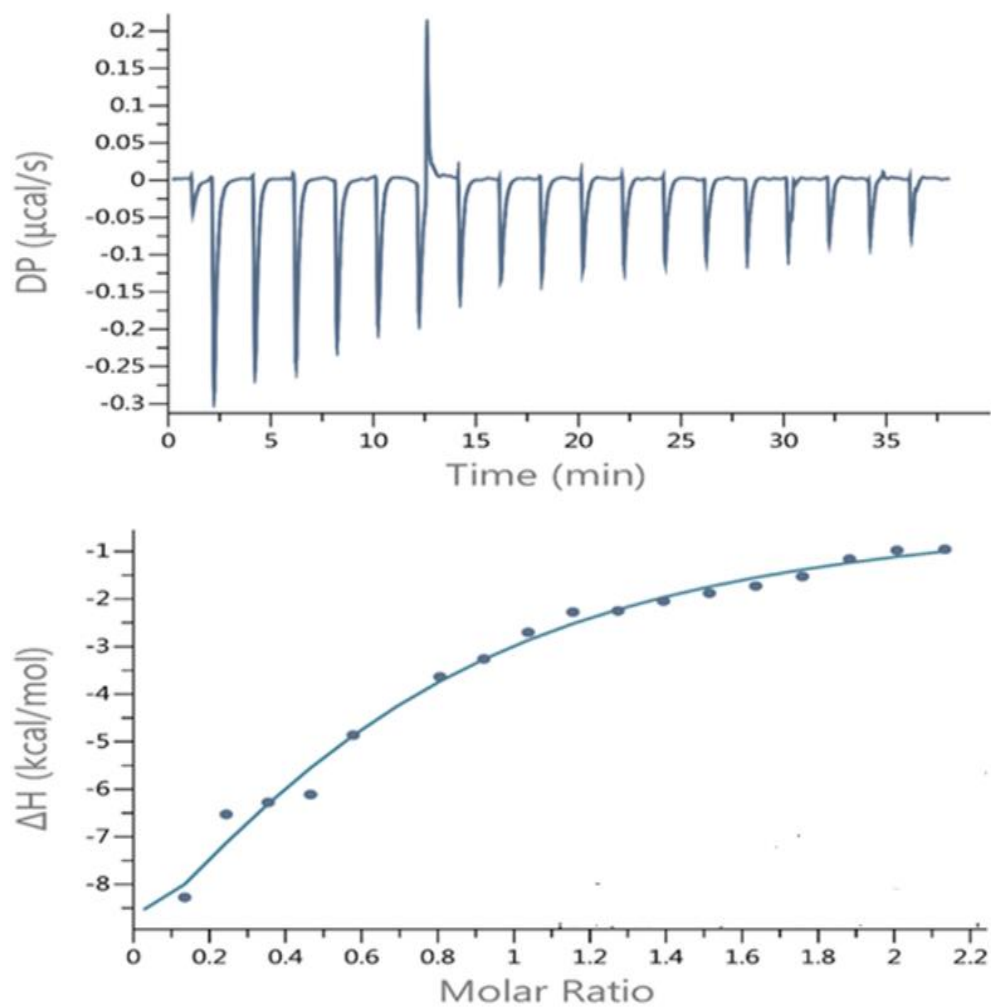


Figure 83. The interaction between USP11 residues (1-244) and Ubpep11S3S4. 200 μM Ubpep11S3S4 was injected into 20 μM USP11 residues (1-244) into the sample cell. The ITC graph results showed a binding curve with an estimated K_d value of about 12.2 μM .

7.4 Discussion of results of Ubiquitin modifications

This chapter has showed the suitability of ubiquitin as a fusion tag and scaffold due to its specific properties being high stability and solubility. Ubiquitin protein can be engineered for high affinity target binding in line with previous studies (269).

The first application of Ubiquitin as an affinity tag was designed by insertion of Histidine tag in the loop between strands 3 and 4. This protein(UbHisS3S4) was expressed and purified successfully. Stability wasn't affected by insertion of the Histidine tag in the loop between strands 3 and 4 as Thermoflower assay results indicated.

On the other hand, UbHisS3S4 was fused with USP20ZNFUBP (1-108) and BBS5(1-270) proteins, these proteins were weakly expressed. The fusion of UBHisS3S4 with BBS5 showed the protein did not bind tightly to nickel column during purification and the protein aggregated when using gel filtration as it was eluted in the void volume. This result may be due to the large molecular weight of the BBS5 affecting the binding affinity of the Histidine tag with the nickel column. This may be due to the position of the Histidine tag in the loop between strands 3 and 4 as the position of the Histidine tag can affect the purification of fusion proteins (275). Also this position of Histidine tag may explain the weak expression of USP20ZNFUBP (1-108) protein. Therefore, UBHisS3S4 has a weak efficiency as an affinity and solubility tag based on these two model proteins.

The second application of Ubiquitin protein was as a protein scaffold to display constrained peptides. This application was investigated and confirmed in a binding experiment by isothermal titration calorimetry. The binding result is promising as it

showed an interaction of Ubpep11S3S4 with the deubiquitinating enzyme USP11 domains residues 1-244. However, this experiment needs additional controls to compare the interaction of Ubpep11S3S4 protein with different proteins to assess its efficiency as a scaffold. Overall, the results show that the insertion of a relatively long peptide (nine amino acid residues) into the loop between strands 3 and 4 of ubiquitin is possible without affecting the fold and suggests that ubiquitin is a valuable protein scaffold and can be used to generate multi-specific molecules for specific therapeutic applications (276).

8. Chapter 8: Conclusion

The first aim of this work was to design and investigate two novel tags. These tags were chosen based on specific criteria from the Protein Data Bank (PDB) (Simon Caulton, PhD thesis, UoN). These criteria were based on the hypothesis that the tags should be small, rigid and globular proteins. Also these tags must have no enzymatic activity (less likely to interfere with a fusion partner sterically or functionally), with reported high solubility and diffracted at high resolution when crystallised. This will aid the expression, solubility of target proteins and its potential use as carrier protein in crystallisation. The 1BKR (13 kDa) and 2GKG tags (14 kDa) were chosen based on the assumption they could promote the expression and solubility of the fusion proteins and also possessed suitable C-termini for fusing to other proteins. The novel tags were expressed and purified at a high yield without any fusion partner which were promising results for our hypothesis to use these tags as solubility tags, as previous studies showed that any protein with such characteristics can be used as a solubility tag. Subsequently, the solubility tag should be expressed and purified in high yield without fusion first and then investigated fused with different model proteins(41).

These new solubility tags have been fused with selected model proteins and their levels of expression were assessed and compared to the activity of a known solubility tag maltose binding protein (MBP). The MBP tag was used in screening comparison as a commonly used as a control with novel tags such as Skp, T7PK and Fh8 tags due to its high expression and purification performance. These tags were compared with MBP tag with different model proteins (41) (211). 1BKR, 2GKG and MBP tags were cloned into the same vector and promoter for consistency and comparative studies (258). All

the proteins were cloned in the pRSF13 vector, varying only in the N terminus fusion partners as any changes in the target protein's expression levels are likely due to the fusion partner's specific features. The small-scale method utilising the qualitative SDS-PAGE was the initial method of comparison between the fused proteins 1BKR, 2GKG and the MBP tag in terms of expression, solubility and purification. The qualitative SDS-PAGE small scale technique has already proven to be a valuable method for the qualitative comparison of various tags and can be used to expect the final yield of protein production in large experiments (245) (277). The small-scale comparisons revealed that the expression level of the 2GKG fusion partners are better than 1BKR fusion partners and are near to the fusion partners of MBP.

The second comparison was carried out in large scale 2 litre bacterial culture for quantitative comparison. Ultimately, the outcomes of the large scale experiments were compatible with the small-scale experiment as 2GKG partners were had a higher yield than 1BKR partners and near to the MBP partners.

All the comparisons included within this study were based on one experiment for each small and large-scale trials and the results from the small-scale trials were largely consistent with what was observed in the large scale trials validating the approach. Previous studies of different novel tags were also based on one experiment such as for Mocr, Fh8, Skp, T7PK and SUMO tags. (41) (68) (2460) (262). These tags were compared with MBP tag which showed the similar results of one of our two novel tags (2GKG) which had almost the level of expression of MBP tag with different model proteins.

The novel solubility tags had quite different impacts on the solubility of the proteins of all selected model proteins. 2GKG generated impressive solubility improvements, almost to the degree of MBP on selected model proteins such as UCHL1, UCHL3, USP20ZNFUBP (1-108) and BBS5(1-270). The 1BKR tag only generated a small enhancement in these proteins solubility compared to the MBP tag. Furthermore, our novel tags were effectively used to express small peptides (less than 100 amino acids) in high yields as peptide production in *E.coli* is difficult as they are prone to degradation.

The novel 1BKR and 2GKG fusions presented in this work largely agreed with our assumption as they were efficient tools for the enhancement of protein solubility and purification in *E. coli*. Thus, these tags can be used in different applications such as protein sample preparation for drug discovery and development of therapeutic agents to treat a wide array of diseases. In addition, this study confirmed that there is no ideal tag for expression of all proteins. Thus, various tags are needed to be tested with the protein of interest and the larger tags generally result in higher production yields of the protein of interest. This is may be due to the large tag slow the translation rate of the fusion protein which enhance the folding of proteins (135). However, in some cases presence of a small tags correlated with high efficiency more than the large tags, as the BBS5 protein results showed the 2GKG tag (14 kDa) was expressed and purified better than that observed with the MBP tag (42.5 kDa). The problem with the larger tags can also be that they are more likely to interfere with a protein's function.

The novel tag crystallization trials failed to crystallize any of the fusion proteins with a fixed arm linker which was not varied and composed of three alanine residues.

On the other hand, these novel tags alone formed crystals. This may be due to many reasons such as:

- 1) The presence of a Histidine tag as affinity tag at the N terminus which can affect the formation of well-ordered diffracting crystals (214).
- 2) The fixed arm linker was not varied and can either be too short or long creating flexibility between the solubility tag and the protein of interest which prevents the formation of crystals (249). This is one parameter that may need optimising for each protein-fusion pair depending on the structures.
- 3) The incubation temperatures (10°C) and (20°C) were not suitable for all proteins to generate crystals as protein crystals generally grow at a preferred temperature which is not known for a new protein. The temperature influences the formation of crystals through affecting the rate of evaporation of the solvent which affects the growth rate of crystals (278).
- 4) The concentrations of protein in the drops may not have been optimal, e.g. too low and was not enough to reach the supersaturation state to form crystals as many drops in the crystallization trays were still clear. The formation of crystals requires large protein concentration as the crystal growth rate increases when protein concentration is high (279).

The second aim of this study was to design and investigate a new fusion tag utilizing the ubiquitin protein by incorporating a His-tag. Previous study showed that ubiquitin protein was used as fusion tag due to its high stability and solubility (280).

This prompted me to investigate whether ubiquitin can be used as affinity and solubility tag to enhance the solubility and stability of a fusion protein when fused to different model proteins. Ubiquitin was engineered by insertion of Histidine tag in the loop between strands 3 and 4. I have chosen to insert the Histidine tag in the loop between the strands 3 and 4 in the ubiquitin protein as the length of this loop varies in ubiquitin fold proteins (281). Previous work also showed that Green Fluorescent Protein (GFP) tag had purified the protein target in a native form with a high purity when a Histidine tag was inserted into the in a loop of the GFP tag (270). The ubiquitin with Histidine tag inserted between the strands 3 and 4 (UBHIS3S4) was expressed and purified alone in a high yield but when fused to two model proteins BBS5(1-270) and USP20 ZNFUBP (1-108) showed a low yield. These results suggested that UBHIS3S4 protein was not favourable for use as affinity and solubility tag or the linker length was not optimal for these cases.

The third aim of this study was to design and investigate the ubiquitin protein as a protein scaffold to be used for high affinity target binding due to its high stability and solubility. This assumption was based on a previous study which showed that the Ubiquitin protein can be engineered to be used for high affinity target binding and is suited as protein scaffold (269). The results of this study of insertion the Histidine tag into the ubiquitin loop between strands 3 and 4 was successful and did not affect the stability and solubility of the ubiquitin protein, which led me to modify this loop of ubiquitin for the interaction studies. Previous research in the Dreveny lab identified a constrained peptide from a constrained phage display library, which interacts with the USP11 N-terminal domains 1-244 (unpublished results). Therefore, the ubiquitin protein was used as a protein scaffold to investigate the binding of a newly identified

peptide with the USP11 residues N-terminal domains (1-244). The interaction of the constrained peptide inserted into the Ubiquitin protein was investigated with USP11 protein residues 1-244 by isothermal titration calorimetry. The results showed binding of USP11 residues 1-244 with by phage display identified constrained peptide, which suggests that ubiquitin can be used as a scaffold in the future to display long constrained peptides.

9. Chapter 9: Future outlook

Further experiments are required to investigate these novel tags to be used as affinity tags as they had not previously been used for affinity purification, so a N-terminal hexa-histidine tag is provided by the cloning vector. MBP and GST tags are frequently used as tags for solubility, affinity and providing various functions for use as fusion partners. These novel tags could possibly be used in this way as the beta spectrin's calponin homology domain (1BKR) binds to EPB41 an erythrocytic cytoskeletal protein (282). Affinity columns might be generated using this protein and a process for purifying 1BKR-tagged proteins might be produced. *FrzS* of *Myxococcus xanthus* PDB(2GKG) tag had no identified ligand until now, so it is unknown if it can be used in the future as an affinity tag.

Further experiments are required to investigate these two novel tags with different expression systems such as mammalian, yeast and insect cells. This will give results for the production of bio-therapeutic proteins as *E. coli* can't express high molecular weight proteins and can't undergo post transitional modification which is required for many therapeutic proteins.

Further studies are also required to design new linker regions between these novel tags and target proteins to be used to crystallise fusion proteins. The design of different linker regions in length and composition can be crucial for the success or failure of the crystallization of fusion proteins (283). Short linkers have been chosen in the linker region between the present novel tags and the protein targets to promote better crystal lattice contacts and to reduce the conformational flexibility of the linker without affecting the structure of the tag and the target protein (66).

In this study the fixed arm linker with three alanine residues was used but it failed to produce any crystals suggesting that the linker may have been too flexible or too rigid (sterically) thus impairing crystal formation. Moon et al investigated different proteins when fused with MBP tag with three alanines (fixed arm linker) but also other linker length which could affect the crystallisation results (284) That's why design of a new linker is required. A linker containing fewer or more alanines instead of three alanine would increase or decrease the flexibility. For example, a penta-alanine linker allowed the interaction between fused protein helping the formation of crystals in the case of the MATa1 protein when fused to MBP tag. Two crystals that diffracted to high resolution were obtained (285).

Further studies are required to use different affinity tags other than Histidine tag with our novel tags. This is because Histidine tag when used with novel tags may be the cause of impairment of crystal formation of model proteins (286). MBP and GST tags can be used as affinity tags which may improve the effect of the novel tags on the expression and purification of the target protein. After purification these affinity tags can be cleaved by specific proteases to avoid any interference of these affinity tags with the target protein crystallization (287). The cleavage of the solubility tags after protein production can also be investigated.

Further studies are required to use these novel tags for crystallization at different incubation temperatures such as 10°C instead of 20°C (used in this work). This may improve crystal growth for specific proteins as the cooler temperature in some cases produced good protein crystals (170).

The position of the Histidine tag, when using ubiquitin protein as affinity and solubility tag, can be changed by insertion of the Histidine tag at either the N or C terminus. The Histidine tag in the loop did not help in purification of the fusion proteins, although this may be due to other reasons. The Histidine tag at N or C terminus can in many cases improve the expression and purification of the target proteins without affecting them (252).

I successfully demonstrated that ubiquitin tolerates long peptide insertions between strands 3 and 4. Further studies can be carried out to use the ubiquitin protein as scaffold with different and even longer peptides to determine the limits and interactions with different proteins measured and compared with the USP11 results. These interaction studies with peptides may lead to create new treatments.

10. References

1. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Molecular cell biology 4th edition. National Center for Biotechnology Information, Bookshelf. 2000.
2. De La Rica R, Matsui H. Applications of peptide and protein-based materials in bionanotechnology. Chemical Society Reviews. 2010;39(9):3499-3509.
3. Ma JK, Chikwamba R, Sparrow P, Fischer R, Mahoney R, Twyman RM. Plant-derived pharmaceuticals—the road forward. Trends in plant science. 2005;10(12):580-585.
4. Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. Biotechnology Advances. 2009;27(3):297-306.
5. Drubin DA, Way JC, Silver PA. Designing biological systems. Genes & development. 2007;21(3):242-254.
6. Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, et al. Protein production and purification. Nature methods. 2008;5(2):135.
7. Chaudhuri TK, Paul S. Protein-misfolding diseases and chaperone-based therapeutic approaches. The FEBS Journal. 2006;273(7):1331-1349.
8. Zhou H-X, Pang X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. Chem Rev. 2018;118(4):1691-1741.
9. Lapidus LJ, Yao S, McGarrity KS, Hertzog DE, Tubman E, Bakajin O. Protein hydrophobic collapse and early folding steps observed in a microfluidic mixer. Biophysical journal. 2007;93(1):218-224.
10. Lopes HS. Evolutionary algorithms for the protein folding problem: A review and current trends. Computational intelligence in biomedicine and bioinformatics: Springer; 2008. p. 297-315.
11. Selkoe DJ. Folding proteins in fatal ways. Nature. 2003;426(6968):900.
12. Martinelli AHS, Lopes FC, John EBO, Carlini CR, Ligabue-Braun R. Modulation of Disordered Proteins with a Focus on Neurodegenerative Diseases and Other Pathologies. Int J Mol Sci. 2019;20(6):1322.

13. Kuang Q, Purhonen P, Hebert H. Two-dimensional crystallization procedure, from protein expression to sample preparation. *BioMed research international*. 2015;2015.
14. Guidi G, Di Tucci L, Santambrogio MD. Profax: A hardware acceleration of a protein folding algorithm. In 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI) 2016 (pp. 1-6). IEEE..
15. Lebendiker M, Danieli T. Production of prone-to-aggregate proteins. *FEBS letters*. 2014;588(2):236-246.
16. Wang W. Protein aggregation and its inhibition in biopharmaceutics. *International journal of pharmaceutics*. 2005;289(1-2):1-30.
17. Frokjaer S, Otzen DE. Protein drug stability: a formulation challenge. *Nature reviews drug discovery*. 2005;4(4):298.
18. Narhi LO, Schmit J, Bechtold-Peters K, Sharma D. Classification of protein aggregates. *Journal of pharmaceutical sciences*. 2012;101(2):493-498.
19. Wang W, Nema S, Teagarden D. Protein aggregation—Pathways and influencing factors. *International Journal of Pharmaceutics*. 2010;390(2):89-99.
20. Mahler H-C, Friess W, Grauschopf U, Kiese S. Protein aggregation: pathways, induction factors and analysis. *Journal of pharmaceutical sciences*. 2009;98(9):2909-2934.
21. Tyedmers J, Mogk A, Bukau B. Cellular strategies for controlling protein aggregation. *Nature reviews Molecular cell biology*. 2010;11(11):777.
22. Sharma VK, Kalonia DS. Temperature- and pH-Induced Multiple Partially Unfolded States of Recombinant Human Interferon- α 2a: Possible Implications in Protein Stability. *Pharmaceutical Research*. 2003;20(11):1721-1729.
23. Wright CF, Teichmann SA, Clarke J, Dobson CM. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature*. 2005;438(7069):878.
24. Zhang W, Xiao S, Ahn DU. Protein oxidation: basic principles and implications for meat quality. *Critical reviews in food science and nutrition*. 2013;53(11):1191-1201.
25. Qiu H, Honey DM, Kingsbury JS, Park A, Boudanova E, Wei RR, et al. Impact of cysteine variants on the structure, activity, and stability of recombinant human α -galactosidase A. *Protein Science*. 2015;24(9):1401-1411.

26. Dahl J-U, Gray MJ, Jakob U. Protein quality control under oxidative stress conditions. *Journal of Molecular Biology*. 2015;427(7):1549-1563.
27. Bhattacharjee S, Dasgupta R, Bagchi A. The Small Heat-shock Proteins IbpA and IbpB in *E. coli*: The Small Ones With a Big Role. *Current Chemical Biology*. 2015;9(2):84-96.
28. Demeule B, Gurny R, Arvinte T. Detection and characterization of protein aggregates by fluorescence microscopy. *International journal of pharmaceutics*. 2007;329(1-2):37-45.
29. Geerlof A, Brown J, Coutard B, Egloff M-P, Enguita F, Fogg M, et al. The impact of protein characterization in structural proteomics. *Acta Crystallographica Section D: Biological Crystallography*. 2006;62(10):1125-1136.
30. Khatun J, Khare SD, Dokholyan NV. Can contact potentials reliably predict stability of proteins? *Journal of molecular biology*. 2004;336(5):1223-1238.
31. Ma B, Nussinov R. Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. *Current opinion in chemical biology*. 2006;10(5):445-452.
32. Borreguero JM, Urbanc B, Lazo ND, Buldyrev SV, Teplow DB, Stanley HE. Folding events in the 21-30 region of amyloid β -protein ($A\beta$) studied in silico. *Proceedings of the National Academy of Sciences*. 2005;102(17):6015-6020.
33. Koschorreck K, Schmid RD, Urlacher VB. Improving the functional expression of a *Bacillus licheniformis* laccase by random and site-directed mutagenesis. *BMC biotechnology*. 2009;9(1):12.
34. Sahdev S, Khattar SK, Saini KS. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and cellular biochemistry*. 2008;307(1-2):249-264.
35. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology*. 2006;72(2):211.
36. Sørensen HP, Mortensen KK. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microbial cell factories*. 2005;4(1):1.
37. Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK. Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microbial cell factories*. 2015;14(1):41.

38. Young TS, Ahmad I, Yin JA, Schultz PG. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *Journal of molecular biology*. 2010;395(2):361-374.
39. Vera A, González-Montalbán N, Arís A, Villaverde A. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and bioengineering*. 2007;96(6):1101-1106.
40. Esposito D, Chatterjee DK. Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology*. 2006;17(4):353-358.
41. Chatterjee DK, Esposito D. Enhanced soluble protein expression using two new fusion tags. *Protein Expression and Purification*. 2006;46(1):122-129.
42. LaVallie ER, DiBlasio-Smith EA, Collins-Racie LA, Lu Z, McCoy JM. Thioredoxin and related proteins as multifunctional fusion tags for soluble expression in *E. coli*. *E coli Gene Expression Protocols*: Springer; 2003. p. 119-140.
43. Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology*. 2003;60(5):523-533.
44. Costa S, Almeida A, Castro A, Domingues L. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Frontiers in microbiology*. 2014;5:63.
45. Fox JD, Kapust RB, Waugh DS. Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. *Protein science : a publication of the Protein Society*. 2001;10(3):622-630.
46. Kwon SB, Ryu K, Son A, Jeong H, Lim K-H, Kim K-H, et al. Conversion of a soluble protein into a potent chaperone in vivo. *Sci Rep*. 2019;9(1):2735.
47. Su Y, Zou Z, Feng S, Zhou P, Cao L. The acidity of protein fusion partners predominantly determines the efficacy to improve the solubility of the target proteins expressed in *Escherichia coli*. *Journal of biotechnology*. 2007;129(3):373-382.
48. Bornhorst JA, Falke JJ. [16] Purification of proteins using polyhistidine affinity tags. *In Methods in enzymology* 2000 Jan 1 (Vol. 326, pp. 245-254).
49. Aricescu AR, Assenberg R, Bill R, Busso D, Chang V, Davis S, et al. Eukaryotic expression: developments for structural proteomics. *Acta Crystallographica Section D: Biological Crystallography*. 2006;62(10):1114-1124.

50. Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnology journal*. 2012;7(5):620-634.
51. Zhao X, Li G, Liang S. Several affinity tags commonly used in chromatographic purification. *Journal of analytical methods in chemistry*. 2013;2013.
52. Harper S, Speicher DW. Purification of proteins fused to glutathione S-transferase. In *Protein Chromatography 2011* (pp. 259-280).
53. Frey S, Görlich D. A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *Journal of chromatography A*. 2014;1337:95-105.
54. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S. Comparison of affinity tags for protein purification. *Protein expression and purification*. 2005;41(1):98-105.
55. Hewitt SN, Choi R, Kelley A, Crowther GJ, Napuli AJ, Van Voorhis WC. Expression of proteins in *Escherichia coli* as fusions with maltose-binding protein to rescue non-expressed targets in a high-throughput protein-expression and purification pipeline. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. 2011;67(9):1006-1009.
56. Butt TR, Edavettal SC, Hall JP, Mattern MR. SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification*. 2005;43(1):1-9.
57. Sachdev D, Chirgwin JM. Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein. *Journal of protein chemistry*. 1999;18(1):127-136.
58. Waugh DS. Making the most of affinity tags. *Trends in Biotechnology*. 2005;23(6):316-320.
59. Xie H, Guo X-M, Chen H. Making the most of fusion tags technology in structural characterization of membrane proteins. *Molecular biotechnology*. 2009;42(2):135-145.
60. Zou Z, Cao L, Zhou P, Su Y, Sun Y, Li W. Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*. *Journal of biotechnology*. 2008;135(4):333-339.
61. Jiang X, Xu J, Yang Q. Soluble expression, purification, and characterization of *Glycydium ussuriensis* venom glosedobin in *Escherichia coli* by using fusion partners. *Applied microbiology and biotechnology*. 2010;85(3):635-642.

62. Yee AA, Savchenko A, Ignachenko A, Lukin J, Xu X, Skarina T, et al. NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *Journal of the American Chemical Society*. 2005;127(47):16512-16517.
63. AhYoung AP, Koehl A, Vizcarra CL, Cascio D, Egea PF. Structure of a putative ClpS N-end rule adaptor protein from the malaria pathogen *Plasmodium falciparum*. *Protein Science*. 2016;25(3):689-701.
64. Shih YP, Kung WM, Chen JC, Yeh CH, Wang AHJ, Wang TF. High-throughput screening of soluble recombinant proteins. *Protein Science*. 2002;11(7):1714-1719.
65. Zhou P, Wagner G. Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies. *Journal of biomolecular NMR*. 2010;46(1):23.
66. Holcomb J, Spellmon N, Zhang Y, Doughan M, Li C, Yang Z. Protein crystallization: Eluding the bottleneck of X-ray crystallography. *AIMS biophysics*. 2017;4(4):557.
67. Malhotra A. Tagging for protein expression. In *Methods in enzymology* 2009 Jan 1 (Vol. 463, pp. 239-258).
68. DelProposto J, Majmudar CY, Smith JL, Brown WC. Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein expression and purification*. 2009;63(1):40-49.
69. Xu X, Song Y, Li Y, Chang J, An L. The tandem affinity purification method: an efficient system for protein complex purification and protein interaction identification. *Protein expression and purification*. 2010;72(2):149-156.
70. Šiurkus J, Neubauer P. Reducing conditions are the key for efficient production of active ribonuclease inhibitor in *Escherichia coli*. *Microbial cell factories*. 2011 Dec;10(1):31.
71. Goh C-S, Milburn D, Gerstein M. Conformational changes associated with protein-protein interactions. *Current opinion in structural biology*. 2004;14(1):104-109.
72. Misura KM, Baker D. Progress and challenges in high-resolution refinement of protein structure models. *Proteins: Structure, Function, and Bioinformatics*. 2005;59(1):15-29.
73. Sinicola JR, Robinson AS. Rapid refolding and polishing of single-chain antibodies from *Escherichia coli* inclusion bodies. *Protein expression and purification*. 2002;26(2):301-308.

74. Lill MA, Danielson ML. Computer-aided drug design platform using PyMOL. *Journal of computer-aided molecular design*. 2011;25(1):13-19.
75. Hammarström M, Woestenenk EA, Hellgren N, Härd T, Berglund H. Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *Journal of structural and functional genomics*. 2006;7(1):1-14.
76. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*. 1999;293(2):321-331.
77. Chan P, Curtis RA, Warwicker J. Soluble expression of proteins correlates with a lack of positively-charged surface. *Sci Rep*. 2013;3:3333.
78. Deller MC, Kong L, Rupp B. Protein stability: a crystallographer's perspective. *Acta Crystallographica Section F: Structural Biology Communications*. 2016;72(2):72-95.
79. Bañuelos S, Saraste M, Carugo KD. Structural comparisons of calponin homology domains: implications for actin binding. *Structure*. 1998;6(11):1419-1431.
80. Fraser JS, Merlie JP, Jr., Echols N, Weisfield SR, Mignot T, Wemmer DE, et al. An atypical receiver domain controls the dynamic polar localization of the *Myxococcus xanthus* social motility protein FrzS. *Molecular microbiology*. 2007;65(2):319-332.
81. Das C, Hoang QQ, Kreinbring CA, Luchansky SJ, Meray RK, Ray SS, et al. Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase UCH-L1. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(12):4675-4680.
82. Johnston SC, Larsen CN, Cook WJ, Wilkinson KD, Hill CP. Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 Å resolution. *The EMBO Journal*. 1997;16(13):3787-3796.
83. Boudreaux DA, Maiti TK, Davies CW, Das C. Ubiquitin vinyl methyl ester binding orients the misaligned active site of the ubiquitin hydrolase UCHL1 into productive conformation. *Proceedings of the National Academy of Sciences*. 2010;107(20):9117-9122.
84. Fischer JA. Deubiquitinating enzymes: their roles in development, differentiation, and disease. *International review of cytology*. 2003:43-72.
85. Maraganore DM, Lesnick TG, Elbaz A, Chartier-Harlin MC, Gasser T, Krüger R, et al. UCHL1 is a Parkinson's disease susceptibility gene. *Annals of neurology*. 2004;55(4):512-521.

86. Wada H, Kito K, Caskey LS, Yeh ET, Kamitani T. Cleavage of the C-terminus of NEDD8 by UCH-L3. *Biochemical and biophysical research communications*. 1998;251(3):688-692.
87. Zhao Q, Xu W, Xing L, Lin Z. Recombinant production of medium-to large-sized peptides in *Escherichia coli* using a cleavable self-aggregating tag. *Microbial cell factories*. 2016;15(1):136.
88. Menzies FM, Huebener J, Renna M, Bonin M, Riess O, Rubinsztein DC. Autophagy induction reduces mutant ataxin-3 levels and toxicity in a mouse model of spinocerebellar ataxia type 3. *Brain*. 2009;133(1):93-104.
89. Matos CA, de Macedo-Ribeiro S, Carvalho AL. Polyglutamine diseases: the special case of ataxin-3 and Machado–Joseph disease. *Progress in neurobiology*. 2011;95(1):26-48.
90. Saunders HM, Bottomley SP. Multi-domain misfolding: understanding the aggregation pathway of polyglutamine proteins. *Protein engineering, design & selection*. 2009;22(8):447-51.
91. Marchut AJ, Hall CK. Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations. *Biophysical journal*. 2006;90(12):4574-4584.
92. Zhemkov VA, Kulminskaya AA, Bezprozvanny IB, Kim M. The 2.2-Angstrom resolution crystal structure of the carboxy-terminal region of ataxin-3. *FEBS open bio*. 2016;6(3):168-178.
93. Xu L, Pu J. Alpha-synuclein in Parkinson's disease: from pathogenetic dysfunction to potential clinical application. *Parkinson's Disease*. 2016;2016.
94. Narhi L, Wood SJ, Steavenson S, Jiang Y, Wu GM, Anafi D, et al. Both familial Parkinson's disease mutations accelerate α -synuclein aggregation. *Journal of Biological Chemistry*. 1999;274(14):9843-9846.
95. Uversky VN. Neuropathology, biochemistry, and biophysics of α -synuclein aggregation. *Journal of neurochemistry*. 2007;103(1):17-37.
96. Zhao M, Cascio D, Sawaya MR, Eisenberg D. Structures of segments of α -synuclein fused to maltose-binding protein suggest intermediate states during amyloid formation. *Protein Science*. 2011;20(6):996-1004.
97. Yuan J, Luo K, Deng M, Li Y, Yin P, Gao B, et al. HERC2-USP20 axis regulates DNA damage checkpoint through Claspin. *Nucleic acids research*. 2014;42(21):13110-13121.

98. Laity JH, Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity. *Current opinion in structural biology*. 2001;11(1):39-46.
99. Bonnet J, Romier C, Tora L, Devys D. Zinc-finger UBPs: regulators of deubiquitylation. *Trends in biochemical sciences*. 2008;33(8):369-375.
100. Khan S, Muhammad N, Khan M, Kamal A, Rehman Z, Khan S. Genetics of human Bardet–Biedl syndrome, an updates. *Clinical genetics*. 2016;90(1):3-15.
101. Hjortshøj TD, Grønskov K, Philp AR, Nishimura DY, Adeyemo A, Rotimi CN, et al. Novel mutations in BBS5 highlight the importance of this gene in non-Caucasian Bardet-Biedl syndrome patients. *American journal of medical genetics Part A*. 2008;146 (4):517.
102. Katsanis N. The oligogenic properties of Bardet–Biedl syndrome. *Human molecular genetics*. 2004;13(suppl_1):R65-R71.
103. Kim Y, Bigelow L, Borovilos M, Dementieva I, Duggan E, Hatzos C, et al. High-throughput protein purification for X-ray crystallography and NMR. *Advances in protein chemistry and structural biology*. 2008;75:85-105.
104. Zampini M, Stevens PR, Pachebat JA, Kingston-Smith A, Mur LAJ, Hayes F. RapGene: a fast and accurate strategy for synthetic gene assembly in *Escherichia coli*. *Scientific reports*. 2015;5:11302.
105. Assenberg R, Delmas O, Graham SC, Verma A, Berrow N, Stuart DI, et al. Expression, purification and crystallization of a lyssavirus matrix (M) protein. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. 2008;64(4):258-262.
106. Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Science*. 2006;15(1):182-189.
107. Lorey S, Fiedler E, Kunert A, Nerkamp J, Lange C, Fiedler M, et al. Novel ubiquitin-derived high affinity binding proteins with tumor targeting properties. *Journal of Biological Chemistry*. 2014;289(12):8493-8507.
108. Schrödel A, Volz J, de Marco A. Fusion tags and chaperone co-expression modulate both the solubility and the inclusion body features of the recombinant CLIPB14 serine protease. *Journal of biotechnology*. 2005;120(1):2-10.
109. Negi H, Reddy PP, Patole C, Das R. A novel polyubiquitin chain linkage formed by viral Ubiquitin prevents cleavage by deubiquitinating enzymes. *bioRxiv*. 2019:756791.

110. Gebauer M, Skerra A. Engineered protein scaffolds as next-generation antibody therapeutics. *Current opinion in chemical biology*. 2009;13(3):245-255.
111. Skerra A. Engineered protein scaffolds for molecular recognition. *Journal of Molecular Recognition*. 2000;13(4):167-187.
112. Spiliotopoulos A, Ferreras LB, Densham RM, Caulton SG, Maddison BC, Morris JR, et al. Discovery of peptide ligands targeting a specific ubiquitin-like domain-binding site in the deubiquitinase USP11. *Journal of Biological Chemistry*. 2019;294(2):424-436.
113. Pal A, Young MA, Donato NJ. Emerging potential of therapeutic targeting of ubiquitin-specific proteases in the treatment of cancer. *Cancer research*. 2014 Sep 15;74(18):4955-4966.
114. Jiang L, Xiao W, Zhou X, Wang W, Fan J. Comparative study of the insoluble and soluble Ulp1 protease constructs as Carrier free and dependent protein immobilizates. *Journal of bioscience and bioengineering*. 2019;127(1):23-29.
115. Kobe B, Ve T, Williams SJ. Fusion-protein-assisted protein crystallization. *Acta Crystallographica Section F: Structural Biology Communications*. 2015;71(7):861-869.
116. Pandhare A, Stuebler AG, Pirayesh E, Jansen M. A modified clear-native polyacrylamide gel electrophoresis technique to investigate the oligomeric state of MBP-5-HT3A-intracellular domain chimeras. *Protein expression and purification*. 2019;153:45-52.
117. Studier FW. Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification*. 2005;41(1):207-234.
118. Chung C, Niemela SL, Miller RH. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences*. 1989;86(7):2172-2175.
119. Van der Rest M, Lange C, Molenaar D. A heat shock following electroporation induces highly efficient transformation of *Corynebacterium glutamicum* with xenogeneic plasmid DNA. *Applied Microbiology and Biotechnology*. 1999;52(4):541-545.
120. Silva LP, Lorenzi PL, Purwaha P, Yong V, Hawke DH, Weinstein JN. Measurement of DNA concentration as a normalization strategy for metabolomic data from adherent cell lines. *Analytical chemistry*. 2013;85(20):9536-9542.
121. Kotchoni SO, Gachomo EW, Betiku E, Shonukan OO. A home made kit for plasmid DNA mini-preparation. *African Journal of Biotechnology*. 2003;2(4).

122. Mullis K, Faloona F, Scharf S, Saiki RK, Horn GT, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In Cold Spring Harbor symposia on quantitative biology 1986 Jan 1 (Vol. 51, pp. 263-273). Cold Spring Harbor Laboratory Press.
123. Ribot EM, Fitzgerald C, Kubota K, Swaminathan B, Barrett TJ. Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*. *Journal of Clinical Microbiology*. 2001;39(5):1889-1894.
124. Elsanhoty RM, Ramadan MF, Jany KD. DNA extraction methods for detecting genetically modified foods: A comparative study. *Food chemistry*. 2011 Jun 15;126(4):1883-1889.
125. Hahn B, Böhm M, Raia V, Zinn N, Möller P, Klingmüller U, et al. One-source peptide/phosphopeptide standards for accurate phosphorylation degree determination. *PROTEOMICS*. 2011;11(3):490-494.
126. Elion EA, Marina P, Yu L. Constructing recombinant DNA molecules by PCR. *Current protocols in molecular biology*. 2007 Apr;78(1):3-17.
127. Bond SR, Naus CC. RF-Cloning. org: an online tool for the design of restriction-free cloning projects. *Nucleic acids research*. 2012;40(W1):W209-W213.
128. Mitchell LA, Cai Y, Taylor M, Noronha AM, Chuang J, Dai L, et al. Multichange isothermal mutagenesis: a new strategy for multiple site-directed mutations in plasmid DNA. *ACS synthetic biology*. 2013;2(8):473-477.
129. Zheng L, Baumann U, Reymond J-L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic acids research*. 2004;32(14):e115.
130. Peabody DS, Manifold-Wheeler B, Medford A, Jordan SK, do Carmo Caldeira J, Chackerian B. Immunogenic display of diverse peptides on virus-like particles of RNA phage MS2. *Journal of molecular biology*. 2008 Jun 27;380(1):252-263.
131. Studier FW. Stable expression clones and auto-induction for protein production in *E. coli*. *Structural Genomics: Springer*; 2014. p. 17-32.
132. Bird LE. High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. *Methods*. 2011;55(1):29-37.
133. Structural Genomics C, Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, et al. Protein production and purification. *Nature Methods*. 2008;5:135.

134. Nguyen H, Martinez B, Oganessian N, Kim R. An automated small-scale protein expression and purification screening provides beneficial information for protein production. *Journal of structural and functional genomics*. 2004;5(1-2):23-27.
135. Berrow NS, Büsow K, Coutard B, Diprose J, Ekberg M, Folkers G, et al. Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. *Acta Crystallographica Section D: Biological Crystallography*. 2006;62(10):1218-1226.
136. Johnson BH, Hecht MH. Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. *Nature Biotechnology*. 1994;12(12):1357.
137. Gaberc-Porekar V, Menart V. Perspectives of immobilized-metal affinity chromatography. *Journal of biochemical and biophysical methods*. 2001;49(1-3):335-360.
138. Harinarayan C, Mueller J, Ljunglöf A, Fahrner R, Van Alstine J, Van Reis R. An exclusion mechanism in ion exchange chromatography. *Biotechnology and bioengineering*. 2006;95(5):775-787.
139. Liu L, Qu M, Liu T, Chen Q, Guo X, Yang J, et al. Biochemical characterization of three midgut chitin deacetylases of the Lepidopteran insect *Bombyx mori*. *Journal of insect physiology*. 2019;113:42-48.
140. Bremm A, Freund SM, Komander D. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nature structural & molecular biology*. 2010;17(8):939.
141. Nishimura DY, Swiderski RE, Searby CC, Berg EM, Ferguson AL, Hennekam R, et al. Comparative genomics and gene expression analysis identifies BBS9, a new Bardet-Biedl syndrome gene. *American journal of human genetics*. 2005;77(6):1021-1033.
142. Arakawa T, Ejima D, Li T, Philo JS. The critical role of mobile phase composition in size exclusion chromatography of protein pharmaceuticals. *Journal of pharmaceutical sciences*. 2010;99(4):1674-1692.
143. Deng J, Davies DR, Wisedchaisri G, Wu M, Hol WG, Mehlin C. An improved protocol for rapid freezing of protein samples for long-term storage. *Acta Crystallographica Section D: Biological Crystallography*. 2004;60(1):203-204.
144. Desjardins P, Hansen JB, Allen M. Microvolume protein concentration determination using the NanoDrop 2000c spectrophotometer. *JoVE (Journal of Visualized Experiments)*. 2009 Nov 4(33):e1610.

145. Fernandez J, Gharahdaghi F, Mische SM. Routine identification of proteins from sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels or polyvinyl difluoride membranes using matrix assisted laser desorption/ionization-time of flight-mass spectrometry (MALDI-TOF-MS). *Electrophoresis*. 1998;19(6):1036-1045.
146. Nettleship JE, Brown J, Groves MR, Geerloff A. Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. *Structural Proteomics: Springer*; 2008. p. 299-318.
147. Boivin S, Kozak S, Meijers R. Optimization of protein purification and characterization using ThermoFluor screens. *Protein expression and purification*. 2013;91(2):192-206.
148. Vollrath F, Hawkins N, Porter D, Holland C, Boulet-Audet M. Differential Scanning Fluorimetry provides high throughput data on silk protein transitions. *Scientific reports*. 2014;4:5625.
149. Mezzasalma TM, Kranz JK, Chan W, Struble GT, Schalk-Hihi C, Deckman IC, et al. Enhancing recombinant protein quality and yield by protein stability profiling. *Journal of biomolecular screening*. 2007;12(3):418-428.
150. Boldon L, Laliberte F, Liu L. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano Rev*. 2015;6:25661.
151. Román-Morales E, López-Alfonzo E, Pietri R, López-Garriga J. Sulfmyoglobin conformational change: A role in the decrease of oxy-myoglobin functionality. *Biochemistry and biophysics reports*. 2016;7:386-393.
152. Valentini E, Kikhney AG, Previtali G, Jeffries CM, Svergun DI. SASBDB, a repository for biological small-angle scattering data. *Nucleic acids research*. 2014;43(D1):D357-D63.
153. Guilbaud J-B, Saiani A. Using small angle scattering (SAS) to structurally characterise peptide and protein self-assembled materials. *Chemical Society Reviews*. 2011;40(3):1200-1210.
154. Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics*. 2007;40(3):191-285.
155. Burger VM, Arenas DJ, Stultz CM. A Structure-free Method for Quantifying Conformational Flexibility in proteins. *Scientific Reports*. 2016;6:29040.

156. Bernado P, Svergun DI. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Molecular biosystems*. 2012;8(1):151-167.
157. Patel TR, Chojnowski G, Koul A, McKenna SA, Bujnicki JM. Structural studies of RNA-protein complexes: a hybrid approach involving hydrodynamics, scattering, and computational methods. *Methods*. 2017;118:146-162.
158. Kavathekar PA, Craig BA, Friedman AM, Bailey-Kellogg C, Balkcom DJ. Characterizing the space of interatomic distance distribution functions consistent with solution scattering data. *Journal of bioinformatics and computational biology*. 2010;8(02):315-335.
159. Grant TD, Luft JR, Carter LG, Matsui T, Weiss TM, Martel A, et al. The accurate assessment of small-angle X-ray scattering data. *Acta Crystallographica Section D: Biological Crystallography*. 2015;71(1):45-56.
160. Shou K, Bremer A, Rindfleisch T, Knox-Brown P, Hirai M, Rekas A, et al. Conformational selection of the intrinsically disordered plant stress protein COR15A in response to solution osmolarity – an X-ray and light scattering study. *Physical Chemistry Chemical Physics*. 2019;21(34):18727-18740.
161. Receveur-Bréchet V, Durand D. How random are intrinsically disordered proteins? A small angle scattering perspective. *Current Protein and Peptide Science*. 2012;13(1):55-75.
162. Mylonas E, Svergun DI. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Applied Crystallography*. 2007;40(s1):s245-s9.
163. Liang Y. Applications of isothermal titration calorimetry in protein science. *Acta biochimica et biophysica Sinica*. 2008;40(7):565-576.
164. Quinn CF, Carpenter MC, Croteau ML, Wilcox DE. Isothermal titration calorimetry measurements of metal ions binding to proteins. In *Methods in enzymology* 2016 Jan 1 (Vol. 567, pp. 3-21).
165. Duff MR, Jr., Grubbs J, Howell EE. Isothermal titration calorimetry for measuring macromolecule-ligand affinity. *J Vis Exp*. 2011(55):2796.
166. Freyer MW, Lewis EA. Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods in cell biology*. 2008;84:79-113.
167. Ball V, Maechling C. Isothermal microcalorimetry to investigate non specific interactions in biophysical chemistry. *Int J Mol Sci*. 2009;10(8):3283-3315.

168. Keller S, Vargas C, Zhao H, Piszczek G, Brautigam CA, Schuck P. High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal Chem*. 2012;84(11):5066-5073.
169. Scheuermann TH, Brautigam CA. High-precision, automated integration of multiple isothermal titration calorimetric thermograms: new features of NITPIC. *Methods (San Diego, Calif)*. 2015;76:87-98.
170. Chun E, Thompson AA, Liu W, Roth CB, Griffith MT, Katritch V, et al. Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure (London, England : 1993)*. 2012;20(6):967-976.
171. Song C, Zhang S, Huang H. Choosing a suitable method for the identification of replication origins in microbial genomes. *Frontiers in microbiology*. 2015 Sep 30;6:1049.
172. Leavitt S, Freire E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Current opinion in structural biology*. 2001;11(5):560-566.
173. Hunte C, Von Jagow G, Schagger H. Membrane protein purification and crystallization: a practical guide: Elsevier; 2003.
174. McPherson A, Gavira JA. Introduction to protein crystallization. *Acta Crystallogr F Struct Biol Commun*. 2014;70(Pt 1):2-20.
175. Simoni RD, Hill RL, Vaughan M. Urease, the first crystalline enzyme and the proof that enzymes are proteins: the work of James B. Sumner. *Journal of Biological Chemistry*. 2002;277(35):e23.
176. Kunitz M. Crystallization of salt-free chymotrypsinogen and chymotrypsin from solution in dilute ethyl alcohol. *The Journal of general physiology*. 1948;32(2):265-269.
177. Summers RG. Present at the Flood: How Structural Molecular Biology Came About (Dickerson, Richard E.). *Journal of Chemical Education*. 2006 Sep;83:1293.
178. Henderson R. Max Perutz (1914–2002): Great Scientist and Modest Leader. *Structure*. 2002;10(4):455-458.
179. Schubert R, Meyer A, Baitan D, Dierks K, Perbandt M, Betzel C. Real-time observation of protein dense liquid cluster evolution during nucleation in protein crystallization. *Crystal Growth & Design*. 2017;17(3):954-958.
180. Zhang X, Zhang P, Wei K, Wang Y, Ma R. The study of continuous membrane crystallization on lysozyme. *Desalination*. 2008;219(1-3):101-117.

181. Bolla JR, Su C-C, Yu EW. Biomolecular membrane protein crystallization. *Philosophical Magazine*. 2012;92(19-21):2648-2661.
182. Schmidt S, Havekost D, Kaiser K, Kauling J, Henzler HJ. Crystallization for the downstream processing of proteins. *Engineering in life sciences*. 2005;5(3):273-276.
183. Benvenuti M, Mangani S. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nature protocols*. 2007;2(7):1633.
184. McPherson A. Protein crystallization. *Protein Crystallography*: Springer; 2017. p. 17-50.
185. Khurshid S, Saridakis E, Govada L, Chayen NE. Porous nucleating agents for protein crystallization. *Nature protocols*. 2014;9(7):1621.
186. Garcia-Ruiz JM. Nucleation of protein crystals. *Journal of Structural Biology*. 2003;142(1):22-31.
187. Maosoongnern S, Diaz Borbon V, Flood AE, Ulrich J. Introducing a fast method to determine the solubility and metastable zone width for proteins: case study lysozyme. *Industrial & Engineering Chemistry Research*. 2012;51(46):15251-15257.
188. McPherson A. A comparison of salts for the crystallization of macromolecules. *Protein Sci*. 2001;10(2):418-422.
189. Krauss IR, Ferraro G, Pica A, Márquez JA, Helliwell JR, Merlino A. Principles and methods used to grow and optimize crystals of protein–metalloidrug adducts, to determine metal binding sites and to assign metal ligands. *Metallomics*. 2017;9(11):1534-1547.
190. Zhang CY, Wu ZQ, Yin DC, Zhou BR, Guo YZ, Lu HM, et al. A strategy for selecting the pH of protein solutions to enhance crystallization. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2013;69(Pt 7):821-826.
191. Li X, Xu X, Dan Y, Zhang M. The factors during protein crystallization: A review. *Crystallography Reports*. 2008;53(7):1261-1266.
192. Iwai W, Yagi D, Ishikawa T, Ohnishi Y, Tanaka I, Niimura N. Crystallization and evaluation of hen egg-white lysozyme crystals for protein pH titration in the crystalline state. *Journal of synchrotron radiation*. 2008;15(3):312-315.
193. Kantardjieff KA, Rupp B. Protein isoelectric point as a predictor for increased crystallization screening efficiency. *Bioinformatics*. 2004;20(14):2162-2168.

194. Judge RA, Jacobs RS, Frazier T, Snell EH, Pusey ML. The effect of temperature and solution pH on the nucleation of tetragonal lysozyme crystals. *Biophys J*. 1999;77(3):1585-1593.
195. Astier J-P, Veesler S. Using temperature to crystallize proteins: a mini-review. *Crystal Growth and Design*. 2008;8(12):4215-4219.
196. Murai R, Nakata S, Kashii M, Adachi H, Niino A, Takano K, et al. Cooling-rate screening system for determining protein crystal growth conditions. *Journal of crystal growth*. 2006;292(2):433-436.
197. McPherson A, Cudney B. Optimization of crystallization conditions for biological macromolecules. *Acta Crystallographica Section F: Structural Biology Communications*. 2014;70(11):1445-1467.
198. Luft JR, Wolfley JR, Said MI, Nagel RM, Lauricella AM, Smith JL, et al. Efficient optimization of crystallization conditions by manipulation of drop volume ratio and temperature. *Protein science*. 2007;16(4):715-722.
199. Majeed S, Ofek G, Belachew A, Huang C-c, Zhou T, Kwong PD. Enhancing protein crystallization through precipitant synergy. *Structure*. 2003;11(9):1061-1070.
200. Potrzebowski W, Trehwella J, Andre I. Bayesian inference of protein conformational ensembles from limited structural data. *PLoS computational biology*. 2018;14(12):e1006641.
201. Baldock P, Mills V, Stewart PS. A comparison of microbatch and vapour diffusion for initial screening of crystallization conditions. *Journal of crystal growth*. 1996;168(1-4):170-174.
202. Heras B, Martin JL. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallographica Section D: Biological Crystallography*. 2005;61(9):1173-1180.
203. Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic acids research*. 2008;36(suppl_2):W5-W9.
204. Hung JH, Weng Z. Sequence alignment and homology search with BLAST and ClustalW. *Cold Spring Harbor Protocols*. 2016 Nov 1;2016(11):pdb-rot093088.

205. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, et al. ExPASy: SIB bioinformatics resource portal. *Nucleic acids research*. 2012;40(W1):W597-W603.
206. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook 2005* (pp. 571-607).
207. Gardès J, Christen R. Design of PCR primers for the detection of waterborne bacteria. *Detection of Pathogens in Water Using Micro and Nano-Technology*. 2012;237.
208. Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A. XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics*. 2007;23(24):3403-3405.
209. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 2006;22(2):195-201.
210. Banuelos S, Saraste M, Djinojic Carugo K. Structural comparisons of calponin homology domains: implications for actin binding. *Structure (London, England : 1993)*. 1998;6(11):1419-1431.
211. Costa SJ, Almeida A, Castro A, Domingues L, Besir H. The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. *Applied microbiology and biotechnology*. 2013;97(15):6779-6791.
212. Sugar FJ, Jenney FE, Poole FL, Brereton PS, Izumi M, Shah C, et al. Comparison of small-and large-scale expression of selected *Pyrococcus furiosus* genes as an aid to high-throughput protein production. *Journal of structural and functional genomics*. 2005;6(2-3):149-158.
213. Liu H, Naismith JH. A simple and efficient expression and purification system using two newly constructed vectors. *Protein expression and purification*. 2009;63(2):102-111.
214. Derewenda ZS. The use of recombinant methods and molecular engineering in protein crystallization. *Methods*. 2004;34(3):354-363.
215. Blommel PG, Becker KJ, Duvnjak P, Fox BG. Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. *Biotechnology progress*. 2007;23(3):585-598.

216. Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein expression and purification*. 2008;59(1):94-102.
217. Rollema HS, Kuipers OP, Both P, De Vos WM, Siezen RJ. Improvement of solubility and stability of the antimicrobial peptide nisin by protein engineering. *Appl Environ Microbiol*. 1995;61(8):2873-2878.
218. Abian O, Grazú V, Hermoso J, González R, García JL, Fernández-Lafuente R, et al. Stabilization of penicillin G acylase from *Escherichia coli*: site-directed mutagenesis of the protein surface to increase multipoint covalent attachment. *Appl Environ Microbiol*. 2004;70(2):1249-1251.
219. Kim I, McKenna SA, Puglisi EV, Puglisi JD. Rapid purification of RNAs using fast performance liquid chromatography (FPLC). *Rna*. 2007;13(2):289-294.
220. Ericsson UB, Hallberg BM, DeTitta GT, Dekker N, Nordlund P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical biochemistry*. 2006;357(2):289-298.
221. Dyson MR, Shadbolt SP, Vincent KJ, Perera RL, McCafferty J. Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC biotechnology*. 2004;4(1):32.
222. Perchiacca JM, Ladiwala ARA, Bhattacharya M, Tessier PM. Aggregation-resistant domain antibodies engineered with charged mutations near the edges of the complementarity-determining regions. *Protein Engineering, Design and Selection*. 2012;25(10):591-602.
223. Simeonov P, Berger-Hoffmann R, Hoffmann R, Sträter N, Zuchner T. Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield. *Protein Engineering, Design and Selection*. 2010;24(3):261-268.
224. Jin T, Chuenchor W, Jiang J, Cheng J, Li Y, Fang K, et al. Design of an expression system to enhance MBP-mediated crystallization. *Scientific reports*. 2017;7:40991.
225. Collins PM, Ng JT, Talon R, Nekrosiute K, Krojer T, Douangamath A, et al. Gentle, fast and effective crystal soaking by acoustic dispensing. *Acta Crystallographica Section D: Structural Biology*. 2017;73(3):246-255.
226. Nasab MA, Javaran MJ, Cusido RM, Palazon J. Purification of recombinant tissue plasminogen activator (rtPA) protein from transplastomic tobacco plants. *Plant physiology and biochemistry*. 2016;108:139-144.

227. Valadares NF, Garratt RC. Septin crystallization for structural analysis. *In*Methods in cell biology 2016 Jan 1 (Vol. 136, pp. 321-338).
228. McKenna SA, Kim I, Puglisi EV, Lindhout DA, Aitken CE, Marshall RA, et al. Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nature protocols*. 2007;2(12):3270.
229. Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. *Current opinion in structural biology*. 2008;18(5):581-586.
230. Gordon SM, Deng J, Lu LJ, Davidson WS. Proteomic characterization of human plasma high density lipoprotein fractionated by gel filtration chromatography. *Journal of proteome research*. 2010;9(10):5239-5249.
231. Chayen NE, Saridakis E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nature Methods*. 2008;5:147.
232. Fee CJ. Size comparison between proteins PEGylated with branched and linear poly (ethylene glycol) molecules. *Biotechnology and bioengineering*. 2007;98(4):725-731.
233. Grant TD, Luft JR, Wolfley JR, Tsuruta H, Martel A, Montelione GT, et al. Small angle X-ray scattering as a complementary tool for high-throughput structural studies. *Biopolymers*. 2011;95(8):517-30.
234. Mertens HD, Svergun DI. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of structural biology*. 2010;172(1):128-141.
235. Kikhney AG, Svergun DI. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters*. 2015;589(19PartA):2570-2577.
236. Kamariah N, Eisenhaber B, Eisenhaber F, Grüber G. Essential role of the flexible linker on the conformational equilibrium of bacterial peroxiredoxin reductase for effective regeneration of peroxiredoxin. *The Journal of biological chemistry*. 2017;292(16):6667-6679.
237. Liu H, Zwart PH. Determining pair distance distribution function from SAXS data using parametric functionals. *Journal of Structural Biology*. 2012;180(1):226-234.
238. Farag AM, Abd-Elnabey HM, Ibrahim HAH, El-Shenawy M. Purification, characterization and antimicrobial activity of chitinase from marine-derived *Aspergillus terreus*. *The Egyptian Journal of Aquatic Research*. 2016;42(2):185-192.

239. Hagel L. Gel-Filtration Chromatography. *Current Protocols in Molecular Biology*. 1998;44(1):10.9.1-9.2.
240. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. 2015;163(3):759-771.
241. Grant TD, Luft JR, Wolfley JR, Tsuruta H, Martel A, Montelione GT, et al. Small angle X-ray scattering as a complementary tool for high-throughput structural studies. *Biopolymers*. 2011;95(8):517-530.
242. Fischer H, de Oliveira Neto M, Napolitano HB, Polikarpov I, Craievich AF. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *Journal of Applied Crystallography*. 2010;43(1):101-109.
243. Durand D, Vivès C, Cannella D, Pérez J, Pebay-Peyroula E, Vachette P, et al. NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of Structural Biology*. 2010;169(1):45-53.
244. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions¹¹ Edited by R. Huber. *Journal of Molecular Biology*. 2000;295(2):337-356.
245. Berrow NS, Bussow K, Coutard B, Diprose J, Ekberg M, Folkers GE, et al. Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. *Acta Crystallographica Section D*. 2006;62(10):1218-1226.
246. Costa SJ, Coelho E, Franco L, Almeida A, Castro A, Domingues L. The Fh8 tag: A fusion partner for simple and cost-effective protein purification in *Escherichia coli*. *Protein Expression and Purification*. 2013;92(2):163-170.
247. Garcia P, Serrano L, Durand D, Rico M, Bruix M. NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation. *Protein science : a publication of the Protein Society*. 2001;10(6):1100-1112.
248. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. Crystal structures of fusion proteins with large-affinity tags. *Protein science : a publication of the Protein Society*. 2003;12(7):1313-1322.
249. Ma B, Kumar S, Tsai C-J, Nussinov R. Folding funnels and binding mechanisms. *Protein Engineering, Design and Selection*. 1999;12(9):713-720.

250. Kim Y, Babnigg G, Jedrzejczak R, Eschenfeldt WH, Li H, Maltseva N, et al. High-throughput protein purification and quality assessment for crystallization. *Methods* (San Diego, Calif). 2011;55(1):12-28.
251. Li Y. Recombinant production of antimicrobial peptides in *Escherichia coli*: A review. *Protein Expression and Purification*. 2011;80(2):260-267.
252. Booth WT, Schlachter CR, Pote S, Ussin N, Mank NJ, Klapper V, et al. Impact of an N-terminal Polyhistidine Tag on Protein Thermal Stability. *ACS omega*. 2018;3(1):760-768.
253. Zhemkov VA, Kulminskaya AA, Bezprozvanny IB, Kim M. The 2.2-Angstrom resolution crystal structure of the carboxy-terminal region of ataxin-3. *FEBS open bio*. 2016;6(3):168-178.
254. Erak M, Bellmann-Sickert K, Els-Heindl S, Beck-Sickinger AG. Peptide chemistry toolbox – Transforming natural peptides into peptide therapeutics. *Bioorganic & Medicinal Chemistry*. 2018;26(10):2759-2765.
255. Lai X, Soler-Lopez M, Wichers HJ, Dijkstra BW. Large-Scale Recombinant Expression and Purification of Human Tyrosinase Suitable for Structural Studies. *PloS one*. 2016;11(8):e0161697.
256. Dale GE, Oefner C, D’Arcy A. The protein as a variable in protein crystallization. *Journal of Structural Biology*. 2003;142(1):88-97.
257. Satakarni M, Curtis R. Production of recombinant peptides as fusions with SUMO. *Protein Expression and Purification*. 2011;78(2):113-119.
258. Nallamsetty S, Waugh DS. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expression and Purification*. 2006;45(1):175-182.
259. Bartels T, Choi JG, Selkoe DJ. α -Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature*. 2011;477(7362):107
260. Malhotra A. Tagging for protein expression. In *Methods in enzymology* 2009 Jan 1 (Vol. 463, pp. 239-258).
261. Kobe B, Ve T, Williams SJ. Fusion-protein-assisted protein crystallization. *Acta crystallographica Section F, Structural biology communications*. 2015;71(Pt 7):861-869.

262. Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci.* 2006;15(1):182-189.
263. Routzahn KM, Waugh DS. Differential effects of supplementary affinity tags on the solubility of MBP fusion proteins. *Journal of structural and functional genomics.* 2002 Jun 1;2(2):83-92.
264. Galkin O, Vekilov PG. Control of protein crystal nucleation around the metastable liquid-liquid phase boundary. *Proceedings of the National Academy of Sciences of the United States of America.* 2000;97(12):6277-6281.
265. McPherson A, Gavira JA. Introduction to protein crystallization. *Acta crystallographica Section F, Structural biology communications.* 2013;70(Pt 1):2-20.
266. Overton IM, van Niekerk CAJ, Barton GJ. XANNpred: neural nets that predict the propensity of a protein to yield diffraction-quality crystals. *Proteins.* 2011;79(4):1027-1033.
267. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research.* 2014;42(Web Server issue):W252-W258.
268. Tokunaga H, Arakawa T, Tokunaga M. Novel soluble expression technologies derived from unique properties of halophilic proteins. *Applied microbiology and biotechnology.* 2010;88(6):1223-1231.
269. Lorey S, Fiedler E, Kunert A, Nerkamp J, Lange C, Fiedler M, et al. Novel ubiquitin-derived high affinity binding proteins with tumor targeting properties. *The Journal of biological chemistry.* 2014;289(12):8493-8507.
270. Kobayashi T, Morone N, Kashiya T, Oyamada H, Kurebayashi N, Murayama T. Engineering a novel multifunctional green fluorescent protein tag for a wide variety of protein research. *PLoS One.* 2008 Dec 2;3(12):e3822.
271. Piana S, Lindorff-Larsen K, Shaw DE. Atomic-level description of ubiquitin folding. *Proceedings of the National Academy of Sciences.* 2013 Apr 9;110(15):5915-5020.
272. Schlaeppi J-M, Henke M, Mahnke M, Hartmann S, Schmitz R, Pouliquen Y, et al. A semi-automated large-scale process for the production of recombinant tagged proteins in the Baculovirus expression system. *Protein Expression and Purification.* 2006;50(2):185-195.

273. Choe W, Durgannavar TA, Chung SJ. Fc-binding ligands of immunoglobulin G: an overview of high affinity proteins and peptides. *Materials*. 2016;9(12):994.
274. Krainer G, Keller S. Single-experiment displacement assay for quantifying high-affinity binding by isothermal titration calorimetry. *Methods*. 2015;76:116-23.
275. Thielges MC, Chung JK, Axup JY, Fayer MD. Influence of histidine tag attachment on picosecond protein dynamics. *Biochemistry*. 2011;50(25):5799-5805.
276. Job F, Settele F, Lorey S, Rundfeldt C, Baumann L, Beck-Sickinger AG, et al. Ubiquitin is a versatile scaffold protein for the generation of molecules with de novo binding and advantageous drug-like properties. *FEBS open bio*. 2015;5:579-593.
277. Dümmler A, Lawrence AM, De Marco A. Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microbial cell factories*. 2005 Dec;4(1):34.
278. Chen RQ, Lu QQ, Cheng QD, Ao LB, Zhang CY, Hou H, Liu YM, Li DW, Yin DC. An ignored variable: solution preparation temperature in protein crystallization. *Scientific reports*. 2015 Jan 19;5:7797.
279. McPherson A, Cudney B. Optimization of crystallization conditions for biological macromolecules. *Acta Crystallogr F Struct Biol Commun*. 2014;70(Pt 11):1445-1467.
280. Paal M, Heel T, Schneider R, Auer B. A novel Ecotin-Ubiquitin-Tag (ECUT) for efficient, soluble peptide production in the periplasm of *Escherichia coli*. *Microbial cell factories*. 2009;8(1):7.
281. Dreveny I, Kondo H, Uchiyama K, Shaw A, Zhang X, Freemont PS. Structural basis of the interaction between the AAA ATPase p97/VCP and its adaptor protein p47. *The EMBO journal*. 2004;23(5):1030-1039.
282. An X, Debnath G, Guo X, Liu S, Lux SE, Baines A, et al. Identification and Functional Characterization of Protein 4.1R and Actin-Binding Sites in Erythrocyte β Spectrin: Regulation of the Interactions by Phosphatidylinositol-4,5-bisphosphate. *Biochemistry*. 2005;44(31):10681-10688.
283. Corsini L, Hothorn M, Scheffzek K, Sattler M, Stier G. Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein science : a publication of the Protein Society*. 2008;17(12):2070-2079.
284. Moon AF, Mueller GA, Zhong X, Pedersen LC. A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci*. 2010;19(5):901-913.

285. Waugh DS. Crystal structures of MBP fusion proteins. *Protein Science*. 2016;25(3):559-571.
286. Kim Y, Babnigg G, Jedrzejczak R, Eschenfeldt WH, Li H, Maltseva N, et al. High-throughput protein purification and quality assessment for crystallization. *Methods*. 2011;55(1):12-28.
287. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. Crystal structures of fusion proteins with large-affinity tags. *Protein Science*. 2003;12(7):1313-1322.