

STRUCTURAL AND FUNCTIONAL STUDIES ON CHFR AUTOUBIQUITINATION

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Leanna Louise Smith BSc (Hons) (Keele)

Department of Molecular and Cell Biology

University of Leicester

September 2017

Structural and functional studies on CHFR autoubiquitination

Leanna Louise Smith

The antephasic checkpoint plays an important role in delaying eukaryotic cell division in the presence of numerous stress conditions. Checkpoint with Fork-Head Associated (FHA) and RING domain (CHFR) is an E3 ubiquitin ligase and an integral component of the antephasic checkpoint. Responsible for delaying mitotic entry in the presence of adverse conditions, downregulation of CHFR is frequently observed in numerous cancer cell lines and tumours. Previous studies have demonstrated N-terminal FHA-domain and C-terminal cysteine rich domain deficient CHFR proteins (CRD; Δ FHA- Δ CRD-CHFR) can form polyubiquitin chains *in vitro*. However, the oligomeric state of the full-length (FL-CHFR) and N-terminal FHA-domain deletion mutant (Δ FHA-CHFR) proteins remains unexplored. This study has demonstrated that the FL-CHFR protein is dimeric in solution, with Δ FHA-CHFR proteins retaining both the capacity to dimerize in solution and form polyubiquitin chains. With di-, tetra- and octomeric CRD species identified, both the FHA domain and the CRD therefore mediate dimerization of FL-CHFR. SEC in-line with Small Angle X-ray Scattering (SEC-SAXS) has verified segment-swapped dimerization of the FHA domain (13-180) in solution, previously observed within a published crystal structure (PDB: 1LGP, 13-125). A screen of 34 E2s has identified 5 previously unreported ubiquitin conjugating enzymes (UbcH5C, UbcH5D, UbcH6, UbcH8, Ubc1) responsible for CHFR mediated polyubiquitin chain formation *in vitro*, corroborating with phylogenetic analysis. A RING domain homology model was generated using X-ray structures of RING homologues, featuring an alpha helix, three antiparallel beta strands and two zinc metal ions; with molecular dynamic simulations revealing considerable Root Mean Square Fluctuations (RMSFs) within the N and C-terminal loop regions. By modelling key interactions within the RING: Ubc13 (:Mms2) ~ ubiquitin complex, site directed FL-CHFR (I306A, P340A, W332A, R345A, H322C/ Y362A, R335A, R343A, E300A, H322C, Y362A) and ubiquitin (E34A, Q40A, R72A, G35A) mutants have identified essential interactions responsible for CHFR-mediated polyubiquitin chain formation *in vitro*.

Acknowledgements

Firstly, I would like to thank my supervisors, Dr Ralf Schmid and Professor Richard Bayliss for their continued guidance and encouragement. Many thanks also to committee members: Dr Raj Patel, Professor Andrew Fry and Dr Aude Echaliier-Glazer for invaluable feedback on my research.

Thank you also to members of the Bayliss group – past and present – for excellent feedback and advice throughout. I owe a special thanks to the members of the Zeqiraj group for making me feel so welcome.

Thank you to the Biotechnology and Biological Sciences Research Council (BBSRC) for funding and to the Midlands Integrative Biosciences Training Partnership (MIBTP) for delivering such a fantastic doctoral training programme.

Finally, I would like to thank my family and friends for their invaluable support and encouragement throughout.

Contents

Abstract	II
Acknowledgements	III
List of tables	X
List of figures	XI
List of Abbreviations	XVII
Chapter 1: Introduction	1
1.1 Eukaryotic cell division and cell cycle checkpoints	2
1.1.1 The antephasis checkpoint	4
1.2 Ubiquitination	5
1.2.1 The ubiquitination pathway	5
1.2.2 E1 ubiquitin activating enzymes	7
1.2.3 E2 ubiquitin conjugating enzymes	7
1.2.4 Mono-, multi- and polyubiquitin chain formation	10
1.2.5 E3 ubiquitin ligases	12
1.2.5.1 HECT E3 ubiquitin ligases	12
1.2.5.2 RING between RING (RBR)	12
1.2.5.3 RING E3 ubiquitin ligases	13
1.2.6 Deubiquitinases (DUBs)	15
1.3 Structure and function of the CHFR RING E3 ubiquitin ligase	16
1.3.1 CHFR-mediated ubiquitination of mitotic proteins delays G ₂ /M transition	16
1.3.2 A role for CHFR in the maintenance of genomic stability	20
1.3.3 CHFR down-regulation via promoter hypermethylation and mutation in cancer cell lines and tumours	21
1.3.4 Domain architecture	22
1.3.4.1 The C-terminal cysteine rich domain (CRD)	24
1.3.4.2 The N-terminal FHA domain	26
1.3.4.3 Central RING domain	28

1.4 Aims and objectives	31
Chapter 2: biochemical and biophysical characterisation of CHFR	33
2.1 Materials and Methods	35
2.1.1 Antibodies	35
2.1.2 Recombinant protein expression and purification	36
2.1.2.1 Expression of FL-CHFR, CHFR Δ FHA 1 to 3, FHA and CHFR C-terminus proteins	36
2.1.2.2 Purification of FL-CHFR	37
2.1.2.3 SDS-PAGE	40
2.1.2.4 Purification of CHFR Δ FHA proteins	41
2.1.2.5 Purification of the CHFR FHA (13-180) domain	41
2.1.2.6 Purification of the CHFR C-terminal cysteine rich domain (394-664)	44
2.1.2.7 Western blots	47
2.1.3 Estimation of protein molecular weights and oligomerisation states	49
2.1.3.1 Analytical size exclusion chromatography	49
2.1.3.2 Native PAGE analysis and western blots against CHFR proteins	52
2.1.4 Prediction of coiled-coil motifs within the full-length CHFR protein	53
2.1.5 Size exclusion chromatography in-line with small-angle X-ray scattering (SEC-SAXS) of CHFR proteins	55
2.1.5.1 Model independent analysis	58
2.1.5.2 Model based analysis	60
2.1.6 Ubiquitination assays	62
2.1.6.1 Expression and purification of His-tagged ubiquitin conjugating enzymes	62
2.1.6.2 Ubiquitination assays and reaction component controls	65

2.1.6.3 Verification of CHFR stability within ubiquitination assay conditions	66
2.1.6.4 FL-CHFR ubiquitination assays in the presence of either the UbcH5a or Ubc13: Mms2 E2 ubiquitin conjugating enzymes.	66
2.1.6.5 Ubiquitination assays to assess whether the FL-CHFR FHA domain is essential for its activities as an E3 ubiquitin ligase	67
2.1.7 Identification of new E2 enzymes responsible for CHFR-mediated polyubiquitin chain formation	68
2.1.7.1 E2Scan assay	68
2.1.7.2 CHIP (E3 ubiquitin ligase) positive control and buffer (E2 dependency) control	70
2.1.7.3 CDC34/ CDC34B ubiquitination assay using untagged E2 enzymes	70
2.1.8 Phylogenetic analysis of E2 ubiquitin conjugating enzymes used in the E2Scan ubiquitination assay	71
2.1.8.1 Multiple sequence alignment of E2s	71
2.1.8.2 Constructing phylogenetic trees using Maximum Likelihood Estimation (MLE)	72
2.1.8.3 Constructing phylogenetic trees using Bayesian inference	73
2.2 Results	74
2.2.1 Dimerization of FL-CHFR is mediated by the C-terminal cysteine-rich domain and not the N-terminal FHA domain	74
2.2.2 The CHFR C-terminus forms a wide range of higher-order oligomers, whilst only dimeric full-length CHFR protein counterparts are observed	78
2.2.3 CHFR dimerization is not the result of coiled-coil motifs outside the FHA, central RING or C-terminal cysteine rich domains	82

2.2.4 Model-independent analysis of the CHFR FHA domain indicates an extended and dimeric protein in solution	87
2.2.5 Model-based analysis of the CHFR FHA domain indicates the FHA domain in solution shares some characteristics with crystal structure counterpart	94
2.2.6 FL-CHFR increases the rate of K63-linked polyubiquitin chain formation in the presence of the Ubc13: Mms2 E2 heterodimer	97
2.2.7 Δ FHA CHFR dimers, inclusive of a CRD, are active E3 ubiquitin ligases	101
2.2.8 Identification of new E2 ubiquitin conjugating enzymes responsible for CHFR ubiquitination	103
2.2.9 CHFR-specific polyubiquitin chain formation in the presence of the CDC34 and CDC34B E2 enzymes may be T7 tag dependent	106
2.2.10 Phylogenetic analysis of E2 ubiquitin conjugating enzymes exhibits some shared evolutionary relationship between E2s responsible for CHFR-mediated polyubiquitin chain formation	108
2.3 Discussion	111
2.3.1 A model of the dimeric CHFR protein	111
2.3.2 Identification of new E2s responsible for CHFR-mediated polyubiquitin chain formation	113
Chapter 3: Homology modelling of the CHFR RING domain	115
3.1 Materials and Methods	122
3.1.1 Multiple sequence alignments and molecular visualisation	122
3.1.2 Secondary structure prediction	123
3.1.3 Domain boundary prediction	125
3.1.4 Homology modelling	126
3.1.4.1 Basic and multi-template modelling	126
3.1.4.2 Model validation and additional loop refinement	129
3.4.1.3 External evaluation resources	131

3.4.1.4 Control model	132
3.4.1.5 Iterative modelling via extensions of the CHFR RING: E2 ~ Ub interfaces	132
3.1.5 Molecular dynamic simulations	134
3.1.6 Computational resources	143
3.2 Results	143
3.2.1 Comparison of RING isoform domain architecture	143
3.2.2 Multiple sequence alignment of CHFR homologs	144
3.2.3 Template selection for homology modelling and RING domain boundary prediction	146
3.2.4 Homology modelling	155
3.2.4.1 CHFR RING domain	155
3.2.4.2 Model validation using ProCheck and ProSA	160
3.2.4.3 Use of a control model to validate the modelling procedure	161
3.2.4.4 Iterative modelling and extension of the CHFR RING domain	169
3.2.5 Molecular dynamic simulations and evaluation of protein model stability	180
3.3 Discussion	186
3.3.1 Secondary structure prediction via PBD structure content proves more informative, in comparison to JPred and PSI-Pred secondary structure prediction tools	186
3.3.2 Loop conformation and interface residue orientation may be sources of error within the CHFR RING models	187
3.3.3 Modelling the CHFR RING stacking interaction	188
3.3.4 Molecular dynamic (MD) simulations indicate good fold and stability of the RING model backbone, with some N- and C- terminal flexibility	188
Chapter 4: Verification of interactions within the CHFR RING: Ubc13 (: Mms2) ~ ubiquitin complex	190

4.1 Materials and Methods	191
4.1.1 Modelling interactions between the CHFR RING domain, Ubc13 E2 and ubiquitin	191
4.1.2 Site directed mutagenesis of full-length CHFR	192
4.1.2.1 Two-step PCR	192
4.1.2.2 Transformation of <i>E. coli</i> DH5α competent cells	195
4.1.2.3 Purification of plasmid DNA	196
4.1.2.4 Quantification of purified DNA	196
4.1.2.5 DNA sequencing	196
4.1.2.6 Expression and purification of FL-CHFR RING mutant proteins	197
4.1.3 Site directed mutagenesis of ubiquitin	197
4.1.3.1 Two-step PCR	197
4.1.3.2 <i>E. coli</i> DH5α transformation, DNA purification and sequencing	198
4.1.3.3 Expression and purification of His6-ubiquitin	199
4.1.4 Ubiquitination assays using FL-CHFR and ubiquitin site directed mutants	199
4.2 Results	199
4.2.1 interfaces within the CHFR RING: Ubc13 ~ ubiquitin complex	199
4.2.2 The majority of FL-CHFR RING domain mutant proteins are dimeric in solution, with MW comparable to the FL-CHFR WT counterpart	204
4.2.3 Verification of modelled interactions at the CHFR RING: Ubc13 and CHFR RING: ubiquitin interfaces	208
4.3 Discussion	215
Chapter 5: Conclusions	218
Chapter 6: Appendices	223
Chapter 7: Bibliography	275

List of Tables

Table 1.1 Important features of selected known E2 ubiquitin conjugating enzymes (Adapted from Ye and Rape, 2009).	9
Table 1.2 Selected summary of CHFR promoter hypermethylation and down regulation identified in primary cancers and cell lines.	22
Table 2.1 LB-agar antibiotic concentrations for bacterial selection.	36
Table 2.2 Calculated molecular weights and deduced oligomerization states of CHFR proteins.	76
Table 2.3 Data collection and scattering parameters for SEC-SAXS analysis of BSA and CHFR FHA domain proteins.	88
Table 3.1 Summary of PDB crystal structures of E3: E2 and E3: E2 ~ ubiquitin complexes used to predict CHFR RING interfaces.	132
Table 3.2 Summary of metal ion centres used to incorporate the ZAFF metal ion parameters of the CHFR RING domain model 10, prior to MD studies (adapted from Li and Merz, 2015).	135
Table 3.3 CHFR RING domain homologous proteins within the local (PDB derived) database.	147
Table 3.4 Prediction of CHFR RING domain amino acid boundaries using InterProScan (Zdobnov and Apweiler, 2010).	155
Table 3.5 Evaluation of CHFR RING domain homology models produced in MODELLER.	156
Table 3.6 RNF8 homologues within the local (PDB derived) database.	162
Table 3.7 Evaluation of RNF8 homology models produced in MODELLER	164
Table 3.8 Evaluation of CHFR RING domain homology models produced in MODELLER	173
Table 3.9 Summary of MODELLER evaluation data and PROCHECK stereochemical quality assessment of extended, multitemplate CHFR RING domain models.	176

Table 4.1 Thermocycler reaction conditions for the first stage of PCR-mediated site directed mutagenesis of full-length CHFR double-stranded DNA (1-664).	194
Table 4.2 Thermocycler reaction conditions for the second stage of PCR-mediated site directed mutagenesis of full-length CHFR double-stranded DNA (1-664).	195
Table 4.3 Thermocycler reaction conditions for the first stage of PCR-mediated site directed mutagenesis of ubiquitin.	198
Table 4.4 Thermocycler reaction conditions for the second stage of PCR-mediated site directed mutagenesis of ubiquitin.	198
Table 4.5 Calculated molecular weights of CHFR RING domain mutant proteins.	207

List of Figures

Figure 1.1 An overview of eukaryotic cell division (adapted from Alberts <i>et al.</i> , 2008; Chin and Yeong, 2010; Kabir <i>et al.</i> , 2016; Aleem and Arceci, 2015).	3
Figure 1.2 An overview of the ubiquitination enzymatic cascade (adapted from Berg <i>et al.</i> , 2007 and Pickart, 2001).	6
Figure 1.3 Structure of the E2 ubiquitin conjugating enzyme UBC domain (Adapted from Stewart <i>et al.</i> , 2016 and Sheng <i>et al.</i> , 2012).	8
Figure 1.4 The ubiquitin code (Adapted from Ye and Rape, 2009).	11
Figure 1.5 Structural features of RING finger domains within E3 ubiquitin ligases (Adapted from Deshaies and Joazeiro, 2009).	13
Figure 1.6 Diversity in the oligomeric states of active RING E3 ubiquitin ligases.	15
Figure 1.7 Current model of CHFR-mediated delay of G ₂ to M transition by ubiquitination of PLK1 and Aurora A mitotic kinases (Adapted from Sanbhani and Yeong, 2012).	18
Figure 1.8 Domain architecture of the full-length CHFR protein, in comparison to protein paralogs and homologs (adapted from Brooks <i>et al.</i> , 2008 and Oberoi <i>et al.</i> , 2010).	23

Figure 1.9 An X-ray structure of the human CHFR C-terminal cysteine rich domain (CRD) (Adapted from Oberoi <i>et al.</i> , 2010, PDB: 2XP0).	25
Figure 1.10 Monomeric and segment swapped dimeric experimental structures of the CHFR FHA domain (Adapted from Stavridi <i>et al.</i> , 2002).	27
Figure 2.1 Schematic representation of a typical SAXS experimental set-up (adapted from Svergun <i>et al.</i> , 2013 and Skou <i>et al.</i> , 2014).	56
Figure 2.2 Calculation of the molecular weights and oligomerization of various CHFR constructs by size exclusion chromatography.	75
Figure 2.3 Native PAGE analysis of the CHFR C-terminal cysteine rich domain (394-664) reveals higher-order assembly complexes.	78
Figure 2.4 Native PAGE analysis of Strep(II) full-length CHFR protein (1-664) reveals a single dimeric species.	81
Figure 2.5 Prediction of FL-CHFR coiled-coil (CC) structural content using the COILS server (Version 2.2.1; Lupas <i>et al.</i> , 1991) indicates that CHFR is very unlikely to contain any coiled-coils.	83
Figure 2.6 Prediction of FL-CHFR, coiled-coil (CC) structural content using the Paircoil2 server (Version 1.0; McDonnell <i>et al.</i> , 2006; Berger <i>et al.</i> , 2006) (with 21 and 28 residue minimum search window) indicates that CHFR is very unlikely to contain any coiled-coils.	85
Figure 2.7 Comparison of (A) COILS (Lupas <i>et al.</i> , 1991) and (B) Paircoil2 (McDonnell <i>et al.</i> , 2006; Berger <i>et al.</i> , 1995) coiled-coil predictions made using the RNF8 protein (Campbell <i>et al.</i> , 2012) and (C) lysozyme (Shoichet <i>et al.</i> , 1995) experimental structures.	86
Figure 2.8 Experimental scattering pattern from the CHFR FHA domain, with buffer subtraction.	89
Figure 2.9 Experimental scattering pattern from the BSA protein control, with buffer subtraction.	90
Figure 2.10 SAXS of the CHFR FHA domain.	91
Figure 2.11 SAXS of the BSA protein control.	92
Figure 2.12 Kratky plots of (A) BSA control and (B) CHFR FHA scattering intensity data.	93
Figure 2.13 Model of the solution structure of the CHFR FHA domain (13-180) in comparison to experimental structures (Stavridi <i>et al.</i> , 2002).	95

Figure 2.14 Validation of the CHFR ubiquitination assay by omitting reaction components.	97
Figure 2.15 Verification of CHFR stability within buffer and temperature conditions of ubiquitination and E2Scan assay conditions.	98
Figure 2.16 CHFR ubiquitination assays performed in the presence and absence of the E3 ubiquitin ligase with the E2s (A) UbcH5a or (B) Ubc13: Mms2.	99
Figure 2.17 The Δ FHA CHFR dimers (with CRD) remain active as E3 ubiquitin ligases.	102
Figure 2.18 Identification of new E2 ubiquitin conjugating enzymes responsible for CHFR ubiquitination using an E2Scan ubiquitination assay.	104
Figure 2.19 CHFR polyubiquitin chain formation is E2 dependent, with E2Scan assay conditions suitable for CHIP E3-mediated polyubiquitin chain formation.	106
Figure 2.20 CHFR polyubiquitin chain formation in the presence of the CDC34 and CDC34B E2 enzymes is T7-tag dependent.	107
Figure 2.21 Unrooted, phylogenetic trees of E2 enzymes within the E2Scan ubiquitination assay kit.	109
Figure 3.1 Overview of the homology modelling process (Adapted from Sanchez and Sali, 2000).	117
Figure 3.2 Three independent strategies for inclusion of metal atoms into molecular mechanics force fields (adapted from Peters <i>et al.</i> , 2010).	121
Figure 3.3 Multiple sequence alignment between human CHFR (Isoform 1) and other closely related CHFR homologues and RING domain containing proteins.	145
Figure 3.4 Multiple sequence alignment between human CHFR (Isoform 1) and the local database structures.	151
Figure 3.5 Zinc ion geometry centres of protein homologues used for basic (single template, PDB: 4AYC) and multi-template (PDBs: 3ZNI, 4AYC, 5D0I, 3NG2, 4S3O and 4QPL) homology modelling of the CHFR RING domain.	154

Figure 3.6 DOPE score profile of the CHFR RING model and template (PDB: 4AYC) obtained in basic, multi-template and loop refinement modelling strategies in MODELLER.	156
Figure 3.7 DOPE score profile of the CHFR RING models loop-refined between alignment positions 6 to 16 in MODELLER.	157
Figure 3.8 Superimposition of the top 5 calculated loop conformations of the CHFR RING domain, refined in MODELLER using the multi-template derived model 39.	158
Figure 3.9 Multiple sequence and cartoon alignment between template (PDB: 4AYC, RNF8) final model (106) of the CHFR RING domain, produced by homology modelling using MODELLER.	159
Figure 3.10 Multiple sequence alignment between RNF8 and the local database structures.	163
Figure 3.11 DOPE score profile of the RNF8 model and template (PDB: 2Y43) obtained in basic, multi-template and loop refinement modelling strategies in MODELLER.	165
Figure 3.12 (A) Superimposition of the top 5 calculated loop conformations of the RNF8 homology models, refined in MODELLER using the multi-template derived model 57.	166
Figure 3.13 Cartoon alignment between PDB: 2Y43 (basic modelling template) and model 85, produced by homology modelling using MODELLER.	167
Figure 3.14 Alignment between RNF8 RING model (641) and PDB deposited structure (4AYC).	168
Figure 3.15 Alignment between the final CHFR RING model (106) and the E3 ubiquitin ligase homologue RNF4 (PDB: 4AP4, Plechanovova <i>et al.</i> , 2012).	170
Figure 3.16 Alignment between Model 106 of the CHFR RING domain with RNF 4 (PDB: 4AP4, Plechanovova <i>et al.</i> , 2012).	172
Figure 3.17 DOPE score profile of the CHFR RING extended model and template (PDB: 4AYC) obtained in basic, multi-template and loop refinement modelling strategies in MODELLER.	173
Figure 3.18 DOPE score profile of the extended CHFR RING domain models loop-refined between alignment positions 58 to 82 in MODELLER.	174

Figure 3.19 Superimposition of the top 5 calculated loop conformations of the extended CHFR RING domain, refined in MODELLER using the multi-template derived model 37.	175
Figure 3.20 Cartoon alignments between (A) CHFR RING models 106 and 10 and (B) template (RING model 106 and RNF4) and RNF4 (PDB: 4AP4), produced by homology modelling using MODELLER.	177
Figure 3.21 Alignment between RNF8 (PDB: 4ORH, Campbell <i>et al.</i> , 2012) RING model 106 and RING model 10.	178
Figure 3.22 Re-evaluation of the CHFR RING: Ubc13 stacking interaction of (A) model 106, in comparison to (B) model 10.	179
Figure 3.23 Re-evaluation of a potential stacking interaction of CHFR RING model 10 and ubiquitin within the RING: Ubc13 ~ ubiquitin complex.	180
Figure 3.24 CHFR RING model 10 (with zinc ions) solvated and neutralized within a rhombic dodecahedron triclinic unit cell (solvated and neutralized in GROMACS; Van der Spoel <i>et al.</i> , 2005).	181
Figure 3.25 Trajectories for CHFR RING (model 10) simulations with zinc.	183
Figure 3.26 Root mean squared fluctuations (RMSFs) with respect to the CHFR RING model 10 structures, with RMSF values calculated for individual (whole) residues.	185
Figure 4.1 Alignment between the final CHFR RING model (10) and the E3 ubiquitin ligase homologue RNF8 (PDB: 4ORH, Campbell <i>et al.</i> , 2012).	200
Figure 4.2 Structure of the CHFR RING model superimposed upon the E3 homologue RNF8, in complex with polyubiquitin C and Ubc13 (using experimental dataset with PDB: 5AIU, Branigan <i>et al.</i> , 2015).	202
Figure 4.3 Stacking interaction modelled between two individual CHFR RING domains and ubiquitin.	203
Figure 4.4 Chromatograms of SEC purified FL-CHFR RING site directed mutants with molecular weight standards (Superdex 200 16/600; GE Healthcare).	205
Figure 4.5 Mutational analysis of the FL-CHFR: Ubc13 (:Mms2)~ Ub complex using CHFR RING site directed mutants.	209

Figure 4.6 Evaluation of CHFR RING residues, modelled to interact with the Ubc13 protein.	210
Figure 4.7 Evaluation of CHFR RING residues, modelled to interact with the ubiquitin protein.	213
Figure 4.8 Mutational analysis of the FL-CHFR: Ubc13 (: Mms2) ~ Ub complex using ubiquitin site directed mutants.	214
Figure 5.1 Model for the dimerization of CHFR.	221

List of abbreviations

µg	Microgram
µl	Microliter
mL	Millilitre
µM	Micromolar
Å	Angstroms
A ₂₈₀	Absorbance at 280
ATP	Adenosine triphosphate
AUC	Analytical ultracentrifugation
C-	Carboxy
CC	Coiled-coil
Cdc	Cell division cycle
CHFR	CHeckpoint with Forkhead Associated and Really Interesting Gene domain
CRD	Cysteine rich domain
Da	Daltons
Dma1	Defective mitotic arrest 1
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DOPE	Discrete optimized potential energy
DSB	Double stranded break
E1	Ubiquitin activating enzyme
E2	Ubiquitin conjugating enzyme
E3	Ubiquitin ligase
FHA	Forkhead associated
kDa	Kilo Daltons
M	Molar
MCPB	Metal cell parameter builder
MD	Molecular dynamics
mg	Miligram

mM	Millimolar
Mms2	Ubiquitin-conjugating enzyme E2 variant 2/ UBE2V2
MolPdf	Molecular probability density function
MPF	Mitosis promoting factor
MW	Molecular weight
MWS	Molecular weight standard
N-	Amino
PAR	Poly[ADP] ribose
PARP1	Poly[ADP-ribose] polymerase
PDB	Protein data bank
PLK1	Polo-like kinase 1
R_g	Radius of gyration
RING	Really Interesting Gene
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SAXS	Small angle X-ray scattering
SEC-SAXS	Small angle X-ray scattering in line with size exclusion chromatography
SEC	Size exclusion chromatography
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
T_m	Melting temperature

CHAPTER 1: Introduction

1.1 EUKARYOTIC CELL DIVISION AND CELL CYCLE CHECKPOINTS

Eukaryotic cell division is the process during which two genetically identical daughter cells are formed, initially involving DNA duplication and proceeded by chromosomal separation via a series of highly coordinated events (Chin and Yeong, 2010). The process can be separated into the four distinct, sequential phases of: G₁, S, G₂ and M (Figure 1.1) (Do, 1997; Alberts *et al.*, 2008; Chin and Yeong, 2010).

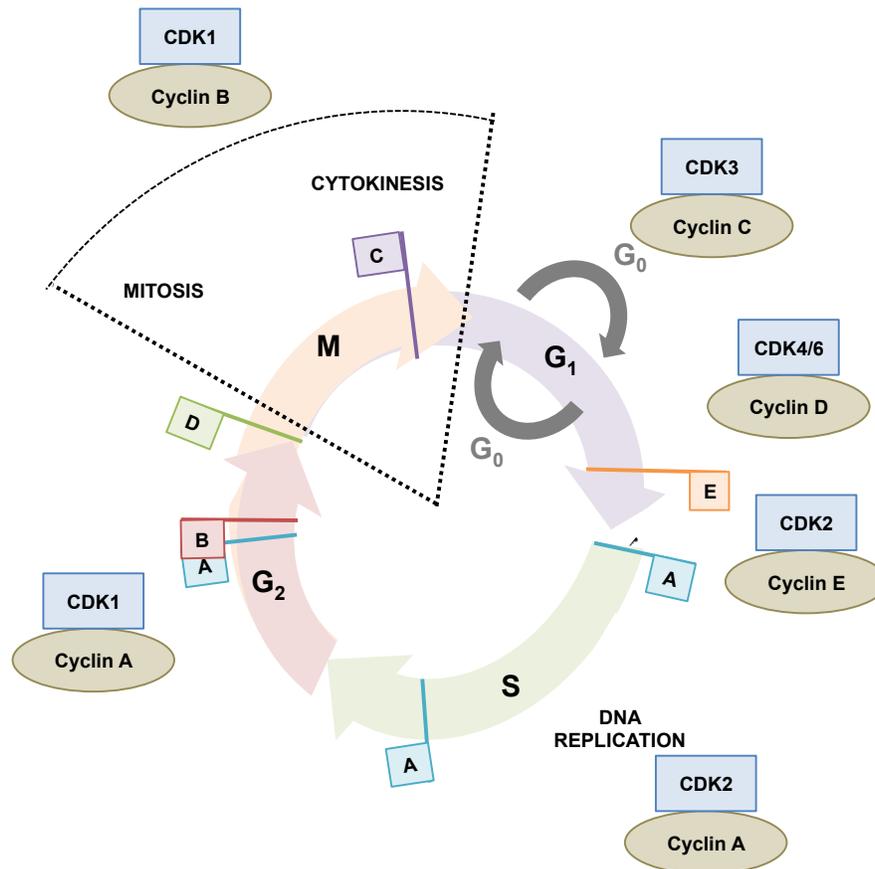


Figure 1.1 An overview of eukaryotic cell division (adapted from Alberts *et al.*, 2008; Chin and Yeong, 2010; Kabir *et al.*, 2016; Aleem and Arceci, 2015). Eukaryotic cell division can be divided into the stages of: G_1 (Gap phase 1), S (DNA Synthesis), G_2 (Gap phase 2) and M (Mitosis) (Do, 1997). G_1 , S and G_2 are collectively known as interphase, occupying 23 out of the total 24 hours of typical cell division (Alberts *et al.*, 2008). Within suboptimal conditions, cells enter G_0 (G zero, or quiescence), during which optimal conditions may be reached (Alberts *et al.*, 2008). Concerning DNA damage, checkpoints occur (A) just prior to entering S-phase at late G_1 , late S phase and halfway during G_2 (indicated by blue boxes) (Chin and Yeong, 2010). Promoting the inheritance of an intact genome, a (B) DNA replication checkpoint partway through G_2 (or just after the final DNA replication checkpoint, pink box) and a (C) Spindle Assembly Checkpoint (SAC, purple box, within mitosis) are also featured (Chin and Yeong, 2010). (D) The antepause checkpoint (green box) delays mitotic entry in conditions exerting cellular stress (Chin and Yeong, 2010). (E) Growth factors mediate the transition of dividing cells from G_1 to S phase, occurring 2 to 3 hours prior to S phase and therefore DNA synthesis (Pardee, 1974; Campisi *et al.*, 1982). At this restriction point (R), growth factors are no longer required for extracellular growth and the cell is now committed to the process of DNA synthesis (Pardee, 1974; Campisi *et al.*, 1982). Complexes formed between cyclin dependent kinase (CDK) and cyclin regulatory subunits are indicated by brown circles and blue boxes, respectively (Aleem and Arceci, 2015; Kabir *et al.*, 2016). Cell cycle entry from G_0 to G_1 is driven by the CDK3-Cyclin C complex; proceeded by CDK4/6-Cyclin D-mediated activation of the downstream CDK2-Cyclin E complex of late G_1 (Aleem and Arceci, 2015; Kabir *et al.*, 2016). Cyclin A regulates S-phase and G_2 (pre-mitotic) progression by forming complexes with CDK2 and CDK1, respectively (Aleem and Arceci, 2015; Kabir *et al.*, 2016). The CDK1-Cyclin B complex is responsible for initiating mitosis (Aleem and Arceci, 2015; Kabir *et al.*, 2016).

Facilitating additional time for eukaryotic cellular growth, both gap phases are also incredibly important to ensuring the external and internal cellular environments are optimum, prior to the substantial commitment to M and S phases (Alberts *et al.*, 2008).

Maintenance of genomic stability is dependent upon the accurate separation of chromosomes between dividing cells (or fidelity), minimizing the probability of chromosomal instability and potentially cellular transformation (Ricke *et al.*, 2008).

1.1.1 The antephasis checkpoint

At the antephasis to prometaphase transition stage of eukaryotic cell division (Bullough and Johnson, 1951), DNA damage, microtubule stress, temperature reduction and damage to the mitotic spindles can initiate the antephasis checkpoint (Chin and Yeong, 2010). Responsible for preventing chromosomal condensation and therefore delaying cell division just prior to commitment to mitosis, antephasis delays the reversible process of chromosomal condensation (Mikhailov *et al.*, 2004, 2005). Components of the antephasis checkpoint include: CHFR (Checkpoint with Forkhead Associated and RING domain); as well activation of the *ataxia telangiectasia mutated* (ATM) (Hari *et al.*, 1995) and phosphatidylinositol kinase-related (ATR) (Cliby *et al.*, 1998) proteins, responsible in mediating a DNA damage response (Hari *et al.*, 1995; Cliby *et al.*, 1998; Bakkenist and Kastan, 2003).

1.2 UBIQUITINATION

1.2.1 The ubiquitination pathway

Ubiquitin is a small (8.5 kDa, 76 residue) protein expressed in all eukaryotic organisms, responsible for targeting proteins for proteasomal degradation via the 26S proteasome within an enzymatic cascade (Hershko and Ciechanover, 1998; Pickart, 2001).

The covalent ligation of the ubiquitin protein or its homologues (for example, SUMO, NEDD8 (Stewart *et al.*, 2016) to intracellular proteins during the process of ubiquitination (or ubiquitylation) is an important reversible post-translational modification, underpinning numerous important biological processes via the degradation of targeted substrates (Hershko and Ciechanover, 1998; Pickart, 2001). The enzymatic, ATP-driven cascade enlists an E1 ubiquitin-activating enzyme, an E2 ubiquitin-conjugating enzyme (or E2 with a pseudo-E2 counterpart), E3 ubiquitin ligase and potentially a substrate protein, targeted for ubiquitination (Hershko and Ciechanover, 1998). Alternatively, the E3 ubiquitin ligase itself can be targeted for autoubiquitination (Pickart, 2001) (Figure 1.2).

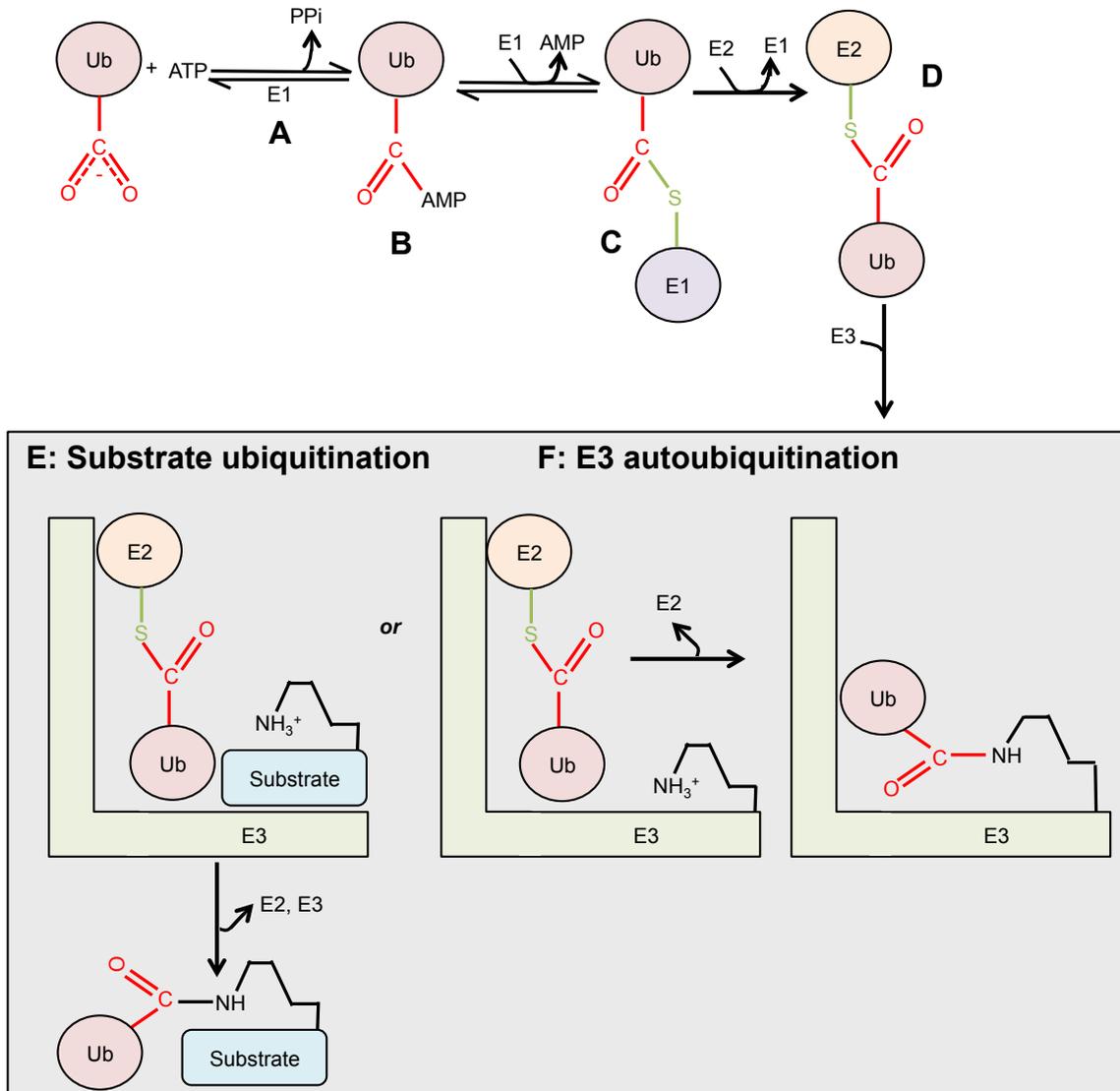


Figure 1.2 An overview of the ubiquitination enzymatic cascade (adapted from Berg *et al.*, 2007 and Pickart, 2001). (A) Within an ATP-driven reaction, ATP (in the presence of a magnesium ion) is converted into AMP and a pyrophosphate is released, (B) with AMP then forming a bond with the C-terminal carboxylate group (G76) of ubiquitin (Pickart, 2001). (C) A sulfhydryl group (of a cysteine residue) within an E1 ubiquitin-activating enzyme adenylates ubiquitin and forms a covalent thioester bond with a (glycine) C-terminal carboxylate of a ubiquitin protein (Pickart, 2001); producing an intermediate ubiquitin adenylate (Hershko and Ciechanover, 1998). (D) Using its di-glycine motif, the E1 then transfers ubiquitin again, this time to the sulfhydryl group of an active site cysteine of an E2 ubiquitin-conjugating enzyme, forming an E2~Ubiquitin thioester linked complex (~ denotes thioester linkage). This is then proceeded by (E) interaction between the E2~ubiquitin complex and an E3 ubiquitin ligase; whereby the C-terminal carboxyl group of ubiquitin forms an isopeptide bond with the ε-amino group of a substrate free terminal amino group. Alternatively, an isopeptide bond can form between C-terminal carboxyl groups of ubiquitin a lysine residue (side chain) (Pickart, 2001). Alternatively, (F) the lysine side chain or ε-amino group of the E3 ubiquitin ligase can be targeted for polyubiquitin chain formation (or autoubiquitination) by the E2-ubiquitin conjugated (Pickart, 2001).

Following ubiquitin chain initiation, long, lysine linked polyubiquitin chains are produced through successive rounds of ubiquitination (or chain elongation) (Pickart, 2001; Beutow and Huang, 2016), mediated by the E3 ubiquitin ligase; linking ubiquitin proteins via one of their seven lysine residues (K6, K11, K27, K29, K33, K48 and K63) (Pickart, 2001). Alternatively, the N-terminal methionine (M1) can also be targeted for polyubiquitin chain formation by the E3 ubiquitin ligase (Hershko and Ciechanover, 1998; Pickart, 2001).

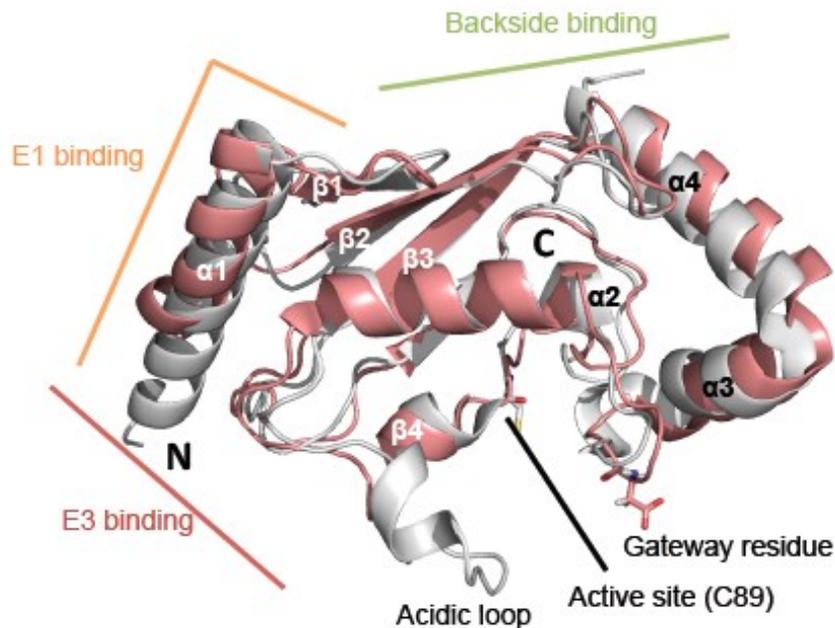
1.2.2 E1 ubiquitin activating enzymes

Within eukaryotes, two E1 ubiquitin activating enzymes are responsible for ubiquitin activation, including: E1 (Accession number: CAA40296.1) and ubiquitin activating enzyme E2 E1 (with 3 isoforms; Accession numbers: NP_003332.1, NP_872607.1 and NP_001189405.1, respectively).

1.2.3 E2 ubiquitin conjugating enzymes

In contrast to the E1 ubiquitin-activating enzyme, a substantially more diverse set of E2 ubiquitin conjugating enzymes confirmed, with at least 40 (inclusive of active and pseudo-E2 counterparts) encoded within the human genome (Sheng *et al.*, 2012). In addition, all E2 ubiquitin conjugating enzymes interact with one of two E1 ubiquitin ligases (Section 1.2.2) (Stewart *et al.*, 2016).

Considerable structural conservation is evident between enzymatically important regions within E2 ubiquitin conjugating enzymes. A 150-residue UBC (catalytic core) domain is observed within all E2 enzymes (containing an α/β -fold) and specific E1 and E3-binding sites identifiable within additional conserved regions (Figure 1.3) (Stewart *et al.*, 2016; Sheng *et al.*, 2012).



UBC7 (E2G2; PDB: 2CYX)

Ubch5c (PDB: 2FUH)

Figure 1.3 Structure of the E2 ubiquitin conjugating enzyme UBC domain (Adapted from Stewart *et al.*, 2016 and Sheng *et al.*, 2012). A cartoon of UBC7 (UBE2G2 or E2G2) grey, PDB: 2CYX; Yoshikawa *et al.*, 2006) is shown superimposed over Ubch5c (PDB: 2FUH; Brzovic *et al.*, 2006). Consisting of four α -helices and four antiparallel β -strands, the conserved α/β -fold is identifiable within different E2 ubiquitin conjugating proteins (Stewart *et al.*, 2016 and Sheng *et al.*, 2012). E3 interactions can be mapped to the N-terminus of first α -helix and a considerable loop region (indicated in red), whilst an E1 binding site is located at the first α -helix (Stewart *et al.*, 2016 and Sheng *et al.*, 2012). E2 activation via phosphorylation of the 'gateway' residue (represented as a stick) exposes the E2 active site and activates its ubiquitin conjugating activities (Yoshikawa *et al.*, 2006; Brzovic *et al.*, 2006). The acidic loop (just outside the E2 catalytic cleft) is also shown for the UBC7 enzyme (Yoshikawa *et al.*, 2006); responsible for interacting with E3 ubiquitin ligases and facilitates the maintenance of an optimal orientation of donor ubiquitin proteins, increasing the accessibility and efficiency of nucleophilic attack via an acceptor ubiquitin (K48) residue (Yoshikawa *et al.*, 2006; Brzovic *et al.*, 2006). Within the catalytic cleft, the active site cysteine for UBC7 (PDB: 2CYX) is indicated by a stick (Yoshikawa *et al.*, 2006). Non-covalent interactions are indicated by the 'backside binding' (green) region, encompassing part of β 1 to 3 and the N-terminus of α -4.

In addition, specific polyubiquitin chain types are produced by distinct E2 ubiquitin conjugating enzymes, with known associated intracellular effects associated with K-linkage type between ubiquitin proteins within the chain (Table 1.1) (Ye and Rape, 2009).

Table 1.1 Important features of selected known E2 ubiquitin conjugating enzymes (Adapted from Ye and Rape, 2009). ‘-’ Denotes unknown.

UniPro accession No.	Name	Other names	Additional features	Chain product	Verified function
P51668	UbcH5A	UBE2D1, E2-17K1	-	K48	26S proteasome
P62837	UbcH5B	UBC4, E2-17K2	-	K48	26S proteasome
P61077	UbcH5C	E2-17K3, UBE2D3	-	K11, K48	-
P51965	UbcH6	UBE2E1,	N-terminal extension	K48	26S proteasome
R68036	UbcH7	E2-F1, L-UBC, UBCE7, UBE2L3	-	K11	26S proteasome
Q96LR5	UbcH8	UBE2E2	N-terminal extension	K48	26S proteasome
Q969T4	UbcH9	UBE2E3, UBCE4, UbcM2, E2-23K	5 isoforms	NEDD8 conjugation	-
O00762	UbcH10	UBE2C, UBXC	N-terminal extension	K11	Cell cycle regulation
P61088	Ubc13	BLU, UBE2N	Forms UBC13-UBE2V1 heterodimer	K63	NFκB signalling, DNA repair
Q15819	MMS2	EDPF-1, UEV2	No active site Cys, UBC13-UEV-2 heterodimer	K63	NFκB signalling, DNA repair
P49427	CDC34	Ubc3, E2-32K,	Acidic loop, C-terminal extension	K48	Cell cycle regulation
Q712K3	CDC34B	Ubc3B, UBE2R2	Acidic loop, C-terminal extension	K48	Cell cycle regulation

1.2.4 Mono-, multi- and polyubiquitin chain formation

The topologies exhibited via the specific types of lysine linkages attached to the target or between individual ubiquitin proteins within a polyubiquitin chain, in addition to the distribution of ubiquitin conjugates covalently linked to the target, underpin the down-stream signalling events initiated by the ubiquitination of either substrate or E3 (Ye and Rape, 2009).

Monoubiquitination, or the addition of a single ubiquitin protein to a target protein (Figure 1.4, Part A), is an important mediator of histone regulation; whereby chromatin remodelling and transcription are underpinned by the covalent attachment of the ubiquitin protein to histones (Wang *et al.*, 2004). For example, the human Polycomb Repressive Complex 1-Like (hPRC1L) complex of multiple E3 ubiquitin ligases (or Polycomb-group) is composed of: HPH2, RING1, RING3 and Bmi1, with monoubiquitination of nucleosomal histone H2A by hPRC1L (at K119) identified and verified within immunoprecipitation assays (Wang *et al.*, 2004).

Multiubiquitination (Figure 1.4, Part B), or the addition of multiple individual and non-linked ubiquitin proteins across a substrate, is frequently attributed to regulation of protein internalization, as first identified by Lee *et al.* (1999). Monoubiquitination of the Colony-Simulating Factor 1 Receptor (CSF-1R)- Cbl complex (CSF-1R: Cbl) is preceded by CSF-1R endocytosis and 26S mediated proteasomal degradation (Lee *et al.*, 1999).

Specific recognition of four or more K11- and K48-linked polyubiquitination chains by the S5a subunit of the 26S proteasome upon intracellular target proteins (Middleton and Day, 2015; Deveraux *et al.*, 1994; Figure 1.4, Part C) results in targeted disassembly of polyubiquitin chains by isopeptidases (for example, isopeptidase T (Wilkinson *et al.*, 1995). A preference of branched K48-linked polyubiquitin chains over linear counterparts has been frequently observed

(Sturner and Behl, 2017; Middleton and Day, 2015), indicative of the specific chain topology underpinning intracellular proteasomal degradation of targeted proteins.

Whilst K48-linked polyubiquitin chains target proteins for proteasomal degradation, mono- and polyubiquitin K63-linked linear counterparts (Figure 1.4, Part D) are instrumental in promoting error-free DNA replication (Chiu *et al.*, 2006) via a proteasomal-independent intracellular pathway (Plans *et al.*, 2006). For example, whilst DNA replication can be achieved via monoubiquitination of the Proliferating Cell Nuclear Antigen (PCNA), K63-linked polyubiquitination promotes the remodelling of inactive DNA replication forks within an alternative error-free pathway (Chiu *et al.*, 2006). In addition, polyubiquitin chain formation with N-terminal (M1) linkages between ubiquitin proteins is targeted exclusively for downstream signalling within the NFκB pathway (Figure 1.4, Part E).

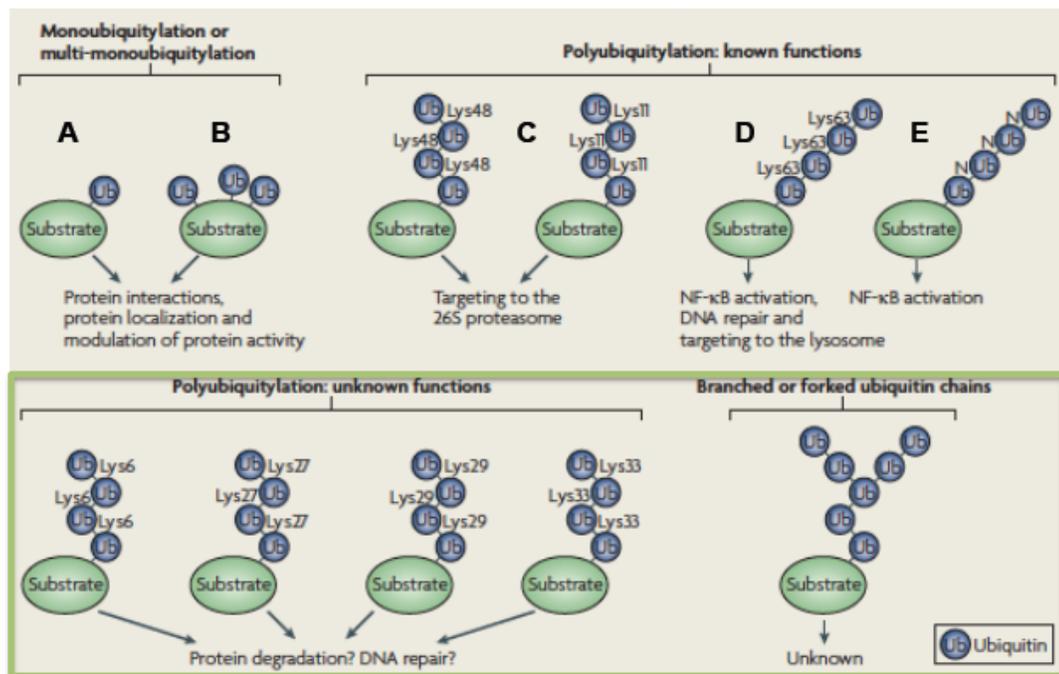


Figure 1.4 The ubiquitin code (Adapted from Ye and Rape, 2009).

Whilst mass spectrometry coupled with *in vitro* and *in vivo* assays has verified the intracellular roles of numerous lysine-specific polyubiquitin chain linkages (Figure 1.4), the role of K6, K27, K29 and K33 -linked polyubiquitin chains remains

unknown; whilst the down-stream effects of branched or forked polyubiquitin chains also remains unknown (Ye and Rape, 2009).

1.2.5 E3 ubiquitin ligases

As the most diverse component of the ubiquitination enzymatic cascade, E3 ubiquitin ligases engage a wide range of molecular mechanisms (Hershko and Ciechanover, 1998). Over 600 different E3 ubiquitin ligases have been identified in mammalian cells (Li *et al.*, 2008) and their classification is dependent upon both mechanism of ubiquitin transfer and the structural basis of E2 recognition (Buetow and Huang, 2016).

1.2.5.1 HECT E3 ubiquitin ligases

One of the major families of eukaryotic ubiquitin ligases (Hershko and Ciechanover, 1998) is the HECT (Homologous to E6-AP Carboxy-terminus) family of E3s. HECT-domain containing E3s catalyse the indirect transfer of ubiquitin to a substrate protein (Hershko and Ciechanover, 1998; 1992) whereby the ubiquitin protein is transferred via two-step reaction; firstly, to the E2 active site cysteine by the HECT E3 ubiquitin ligase (Scheffner *et al.*, 1995) forming an thioester-linked E3~ubiquitin complex (Buetow and Huang, 2016). The ubiquitin protein is then attached to the target protein via an amide bond (Scheffner *et al.*, 1995).

1.2.5.2 RING between RING (RBR)

Consisting of two individual RING domains (RING1 and RING2), RING between RING (RBR) E3 ubiquitin ligases are structurally distinct from other RING E3 ubiquitin ligases; since they also include a third central in-between RING (IBR) domain (Marin and Ferrus, 2002). Ubiquitination of substrates via RBR E3 ubiquitin ligases involves a two-step mechanism. Initially, the E2~ubiquitin thioester conjugate is recruited by RING1, proceeded by transfer of the ubiquitin protein to the RING2 catalytic cysteine (Yuan *et al.*, 2017).

1.2.5.3 RING E3 ubiquitin ligases

In contrast to RBR and HECT E3s, Really INteresting Gene (RING) E3 ubiquitin ligases, RING domain containing counterparts mediate the direct transfer of the ubiquitin protein (within the E2-ubiquitin thioester conjugate) to a substrate, or to the E3 itself (during autoubiquitination) (Metzger *et al.*, 2014; Mattioli and Sixma, 2014). They are considered the second major family of RING E3 ubiquitin ligases in eukaryotes (Hershko and Ciechanover, 1998; 1984).

1.2.5.3.1 Structural features of the RING domain

RING E3 ubiquitin ligases contain a conserved, cysteine-rich zinc binding motif (or RING domain): C-X₂-C-X-(9-39)-X-X(1-2)-H-X(2-3)-X-X₂-X-X(4-48)-C-X₂-X, where X represents any residue (Freemont *et al.*, 1991) (Figure 1.5); essential for both engagement with the E2~ubiquitin thioester conjugate and ubiquitin transfer (to a substrate or itself) (Deshaies and Joazeiro, 2009).

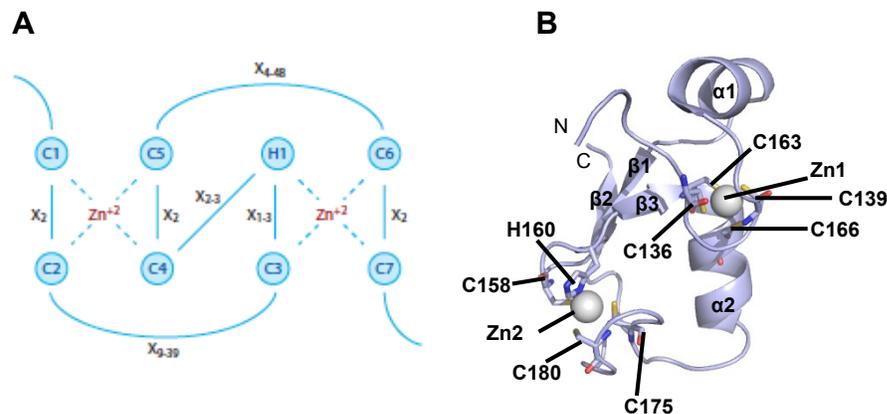


Figure 1.5 Structural features of RING finger domains within E3 ubiquitin ligases (Adapted from Deshaies and Joazeiro, 2009). (A) Schematic of the conserved RING domain sequence, with characteristic cross-brace motif (Metzger *et al.*, 2014). Conserved cysteine and histidine residue positions (blue circles) and coordination to respective first and second zinc ion metal binding centres (shown in red) are also indicated (Deshaies and Joazeiro, 2009; Scheffner *et al.*, 1995). Whilst the zinc ion binding centres are responsible for maintaining correct fold of the CHFR central RING domain, co-ordinating histidine and cysteine residues are important in optimal E3 ubiquitin ligase activity (Scheffner *et al.*, 1995). (B) Cartoon of a single RING finger domain within the RNF4 E3 ubiquitin ligase (PDB: 4AP4, Plechanovova *et al.*, 2012). Grey spheres indicate zinc metal ion binding centres, whilst sticks represent key co-ordinating residues.

1.2.5.3.2 Interaction with E2 ubiquitin conjugating enzymes

E3 RING ubiquitin ligase binding sites have been identified within E2s (Figure 1.3), with a conserved region specifically located at the N-terminus of the first α - helix and loop region of the ubiquitin conjugating enzyme (Sheng *et al.*, 2012; Hershko, and Ciechanover, 1992; Scheffner *et al.*, 1995). Site directed mutagenesis of the E2 binding site within the BRCA1/ BARD1 heterodimeric E3 ubiquitin ligase, in conjunction with NMR, identified that the two loop regions between the zinc metal ion binding centres are essential in E2 recognition by the E3 ubiquitin ligases (Brzovic *et al.*, 2003).

Alternatively, mutations within the central helix of RING domain of E3 ubiquitin ligases can ablate their *in vitro* substrate ubiquitination activities. For example, recognition of tyrosine phosphorylated substrates by the c-Cbl RING E3 ubiquitin ligase and RING finger-specific recruitment of Ub_{CH5a}~ ubiquitin thioester conjugates is ablated upon a Y408A mutation, located within the c-Cbl central alpha helix (Brzovic *et al.*, 2003; Joazeiro *et al.*, 1999).

1.2.5.3.3 Oligomerization of RING E3s

A diverse range of oligomeric states is exhibited and may be required for activation of selected RING E3 ubiquitin ligases.

Firstly, E3s can act as active, monomeric ubiquitin ligases, such as the monomeric Arc-like RING, with equal affinity observed in both dimeric and monomeric oligomerisation states (Wright *et al.*, 2016) (Figure 1.6, Panel A). Secondly, heterodimerization is essential for selected RING E3 ubiquitin ligase activities, as exhibited within the MDM2-MDMX RING E3 heterodimer (Linke *et al.*, 2008) (Figure 1.6, Panel B). Homodimeric E3 ubiquitin ligase counterparts (Figure 1.6, Panel C), such as the RNF8, also require dimerization for activation (Campbell *et al.*, 2012). Alternatively, multicomponent complexes of individual RING E3 ligases

may also be required for full activity, evident in the Anaphase Promoting Complex (APC) (Figure 1.6, Panel D) (Chang *et al.*, 2015).

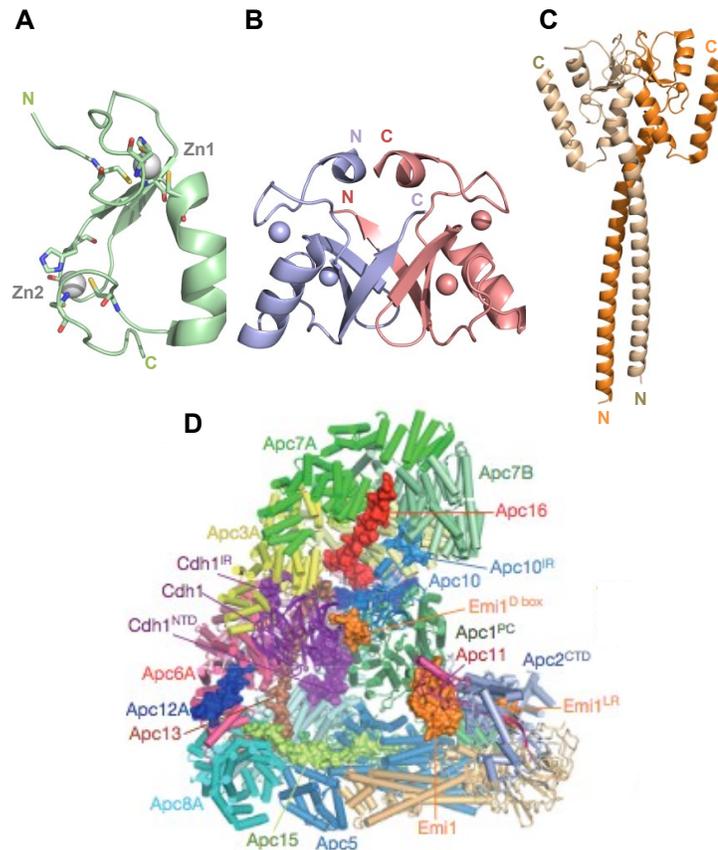


Figure 1.6 Diversity in the oligomeric states of active RING E3 ubiquitin ligases. (A) The active monomeric Arc-like RING (PDB: 5D0I, Wright *et al.*, 2016), (B) heterodimeric MDM2-MDMX (purple and salmon cartoons, respectively; PDB: 2VJE, Linke *et al.*, 2008), (C) homodimeric RNF8 (PDB: 4ORH, Campbell *et al.*, 2012) and (D) multicomponent Anaphase Promoting Complex/ C (APC/C) (Chang *et al.*, 2015; Zhang *et al.*, 2010); exemplify the diversity of RING E3 oligomeric states required for activity. Spheres indicate coordinated zinc ions. Panel D is adapted from Chang *et al.* (2015).

Equally, a hugely diverse range of mechanisms mediate the oligomerization of individual RING E3 proteins, including extensive coiled-coil domains (RNF8, Campbell *et al.*, 2012; Figure 1.6, Panel C) or scaffold formation, as exhibited in the Cullin-RING ligase multisubunit protein complex (Duda *et al.*, 2008).

1.2.6 Deubiquitinases (DUBs)

As a reversible post-translational modification of proteins, ubiquitination of target proteins can be reversed by removal of the ubiquitin protein by deubiquitinases

(DUBs, or deubiquitylating enzymes) (Evans *et al.*, 2003); mediating regulation and homeostasis proceeding the ubiquitination enzymatic cascade and targeted ubiquitination of substrate proteins (Komander, 2010; Komander *et al.*, 2009). With around 79 individual DUBS identified within the human genome (Nijman *et al.*, 2005), approximately sixteen different ubiquitin binding domains have been characterized (Hurley *et al.*, 2006).

1.3 STRUCTURE AND FUNCTION OF THE CHFR RING E3 UBIQUITIN LIGASE

Ubiquitously expressed in all normal tissues, Checkpoint with Forkhead Associated and RING domain (CHFR) is a RING E3 ubiquitin ligase responsible for delaying mitotic progression at the antepause checkpoint in response to cellular stress (Scolnick and Halazonetis, 2000).

The *chfr* gene is located on the reverse strand of Chromosome 12 (132,822,187-132,956,304), with 9 splice variants identified in humans (Aken *et al.*, 2016)

1.3.1 CHFR-mediated ubiquitination of mitotic proteins delays G₂/ M transition

Predominantly located within the nucleus, CHFR contains lysine-rich Nuclear Localisation Motif (NLM; 257-259, KKK), essential in both checkpoint functions and protein recruitment to the nucleus (Kwon *et al.*, 2009).

CHFR is responsible for delaying mitotic progression to prophase (Scolnick and Halazonetis, 2000) at the antepause checkpoint (Matsusaka and Pines, 2004; Scolnick and Halazonetis, 2000) in response to cell stress (Matsusaka and Pines, 2004; Scolnick and Halazonetis, 2000). Cells exhibiting no detectable *chfr* expression progress to metaphase; exhibiting a higher mitotic index, despite administration of a microtubule depolymerizing drugs (nocodazole); in comparison to *chfr* expressing (WT), CHFR (stably) - transfected and vector transfected

(control) counterparts (Scolnick and Halazonetis, 2000). Using a DLD1 (CHFR-deficient) cell line, Scolnick and Halazonetis (2000) demonstrated that following stable transfection with CHFR, entry into metaphase is delayed by approximately 6 hours. These findings were comparable across multiple cell lines and following administration of microtubule-targeting drugs (Scolnick and Halazonetis, 2000; Pines and Matsusaka, 2004). In contrast, delayed mitotic progression was not observed in both CHFR-deficient cell lines and non-transfected controls (Scolnick and Halazonetis, 2000; Pines and Matsusaka, 2004) with nucleoli breakdown also observed (Pines and Matsusaka, 2004) indicating an important role for the protein in presence of intracellular stress. A role for CHFR in maintaining genomic integrity was also explored, whereby poor cell viability, aneuploidy and fragmented nuclei was observed in CHFR-deficient DLD1 cells, in comparison to CHFR-transfected controls (Scolnick and Halazonetis, 2000).

Expression levels of the CHFR protein within cells may be highest at prophase, with very low levels detected at G₁ and a slight elevation at late S phase, in comparison to asynchronous control counterparts (Chaturvedi *et al.*, 2002). However, ascertaining the mitotic-specific stages of CHFR expression level has proved very difficult to reproduce (Bothos *et al.*, 2003), with difficulty potentially exacerbated by differences in cell lines used and high variability between checkpoint assays.

CHFR may play an important role in regulating the intracellular concentrations of important mitotic protein kinases via ubiquitination, proteasomal degradation and/or intracellular signalling; targeting Polo-Like Kinase 1 (PLK1) (Kang *et al.*, 2002) and Aurora A (Yu *et al.*, 2005) as substrates (Figure 1.7).

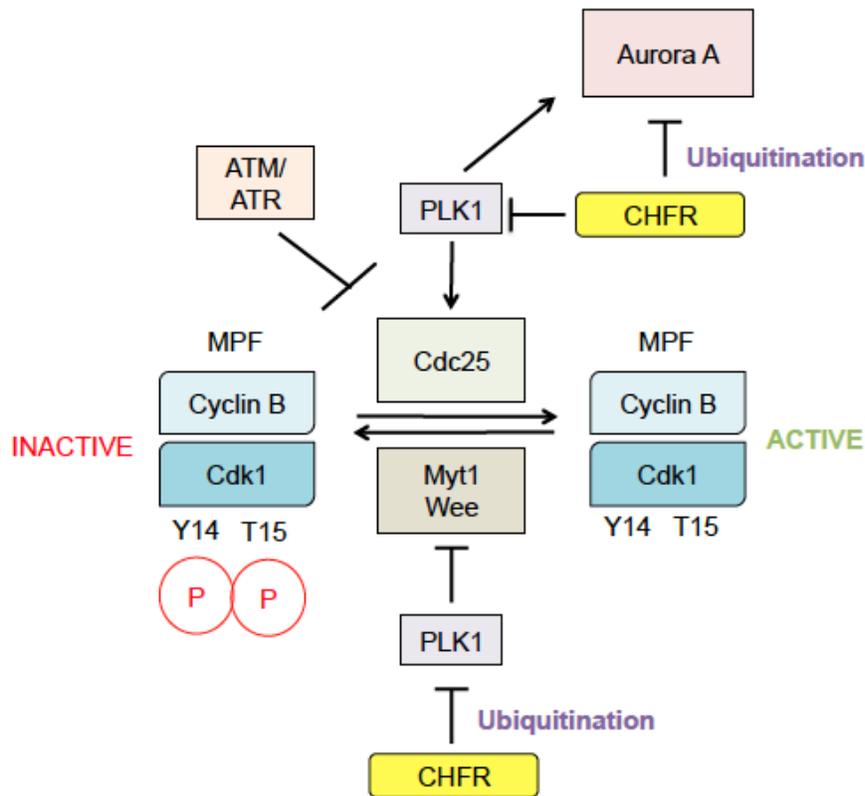


Figure 1.7 Current model of CHFR-mediated delay of G2 to M transition by ubiquitination of PLK1 and Aurora A mitotic kinases (Adapted from Sanbhani and Yeong, 2012). The mitotic promoting factor (MPF) is a heterodimeric complex composed of Cyclin B and Cdk1 (King *et al.*, 1994) and within its active state, responsible for promoting mitotic spindle formation, nuclear envelope breakdown and chromosomal condensation (Kang *et al.*, 2002). Elevated expression and increased intracellular concentrations of the Mitosis Promoting Factor (MPF) complex promotes its localization to the nucleus within prophase (Murray and Kirschner, 1989). Within an active state, both Y14 and T15 of Cdk1 are dephosphorylated by Cdc25 (Russell and Nurse, 1986). However, phosphorylation of Cdk1 (at Y14 and T15) by the Wee and Myt1 mitotic kinases inactivates the MPF complex (Muller *et al.*, 1995, Russell and Nurse, 1986). A third posttranslational modification is responsible for regulating the MPF activities during mitotic division (Fang *et al.*, 1999). At late antepause, ubiquitination of Cyclin B and targeted proteasomal degradation results in deactivation of Cdc25, leaving the MPF susceptible to phosphorylation by Myt1 and Wee, resulting in an inactive MPF complex (Fang *et al.*, 1999). Kang *et al.* (2002) have demonstrated (via addition of recombinant CHFR to *Xenopus* extracts) that CHFR mediates ubiquitination of Polo-Like Kinase 1 (PLK1), resulting in Cdc25 remaining within an inactive state, promoting the phosphorylative (and therefore deactivating) activities of Myt1 and Wee kinases, retaining the phosphorylated (and inactive) state of the MPF complex. CHFR may target Aurora A for proteasomal degradation, whereby overexpression of the mitotic kinase and centrosomal duplication is evident in siRNA *chfr* knockdown cell lines (Privette *et al.*, 2008). Upon detection of DNA double stranded breaks (DSBs), ataxia telangiectasia mutated (ATM) and ataxia telangiectasia mutated- and Rad3 related (ATR) deactivate Cdc25 (Wu *et al.*, 2011; Cillby *et al.*, 1998).

In vitro, ubiquitination of PLK1 is mediated by both the CHFR E3 ubiquitin ligase and Ubc4 E2 ubiquitin-conjugating enzyme (Kang *et al.*, 2002). However, Matsusaka and Pines (2004) were unable to detect a reduction in intracellular PLK1 upon activation of the antephasis checkpoint and inhibition of the 26S proteasome has no effect, suggesting Ubc13: Mms2-mediated formation of K63-linked polyubiquitin conjugates (Bothos *et al.*, 2003) on PLK1 (Kang *et al.*, 2002; Matsusaka and Pines, 2004) may target the kinase for a role in intracellular signalling, rather than proteasomal degradation (associated with K48-linked polyubiquitin chains) (Matsusaka and Pines, 2004). However, ubiquitination is essential for function of the antephasis checkpoint, since methyl-ubiquitin injection of Ptk1 cells is typically preceded by mitotic progression, even in the presence of microtubule targeting drugs (Matsusaka and Pines, 2004). It remains unclear as to the role of CHFR in targeting PLK1 for ubiquitination (Matsusaka and Pines, 2004) and the down-stream effects this may have in assigning a role in signalling and delay of mitotic progression at the antephasis checkpoint.

Yu *et al* (2005) demonstrated CHFR's role in regulating *in vivo* concentrations of Aurora A, whereby 5.3 to 9.0 % of CHFR knock-out mice exhibit spontaneous tumour formation within the liver, lungs and GI tract and over 30% of transformed cells exhibit significant chromosomal defects (including aneuploidy and polyploidy). N- and K63 linked polyubiquitin chains and down-regulation of the NFκB pathway via CHFR is attributed to both a significant reduction in tumour vasculature and inhibition of human endothelial cell migration, mediated by the CHFR-dependent downregulation of NFκB and the down-stream loss of IL8 expression, independent of the central CHFR RING domain (Kashima *et al.*, 2009).

Whilst CHFR-mediated ubiquitination of protein kinase substrates (PLK1, Aurora A) have been verified and partially characterized (Kang *et al.*, 2002, Privette *et al.*, 2008; Matsusaka and Pines, 2004), the exact molecular mechanism underpinning the CHFR RING: protein kinase interactions responsible for targeting proteins for proteasomal degradation or intracellular signalling remain unknown. In addition, the

preceding interaction between the CHFR RING domain and the E2~ubiquitin thioester conjugate, prior to substrate ubiquitination (Figure 1.2), remains poorly uncharacterised.

1.3.2 A role for CHFR in the maintenance of genomic stability

CHFR maintains genomic stability by interacting with proteins associated with the mitotic spindle assembly checkpoint, such as MAD2 (Privette *et al.*, 2008). siRNA-mediated knock-downs of *chfr* expression within stable cells line (immortalized human mammary epithelial cells, IHMEC) ablates a key MAD2: CHFR interaction, essential in correct MAD2/ CDC20 localization (Privette *et al.*, 2008). In addition, CHFR-depleted cells are typically aneuploid, polyploid or tetraploid (Privette *et al.*, 2008; Yu *et al.*, 2005), indicative of significant loss in genomic integrity during mitotic cell division.

In addition, Wu *et al* (2011) explored the ubiquitination activities and cell cycle arrest (Figure 1.7) by both CHFR and the RNF8 RING E3 homologue in response to DNA double stranded breaks (DSBs). Upon DSB formation, the ATM and ATR protein kinases become active, targeting down stream proteins (such as BRCA and p53; Kashima *et al.*, 2009) for phosphorylation (Wu *et al.*, 2011; Goodzari *et al.*, 2010).

RNF8 and CHFR, together with the Ubc13 E2 enzyme, impose a synergistic mechanism of DSB-specific ubiquitination cascade of multiple histones (H2A and H2B), promoting chromatin remodelling (via relaxation of chromatin; Liu *et al.*, 2013) and repair of DSBs (Goodzari *et al.*, 2010; Wu *et al.*, 2011). In addition, Poly [ADP-ribose] polymerase 1 (PARP1) can recruit CHFR (via the C-terminal PAR binding domain, or PBZ) and RNF3 to DSB, facilitating chromatin remodelling (Liu *et al.*, 2013). Upon initiation of DSB repairs and PARP-1 upregulation, co-localization and increased intracellular concentration of the CHFR: PARP-1 complex is evident (Kashima *et al.*, 2012). Following the initial E3: PARP-1

complex formation, CHFR ubiquitinates and targets PARP1 for proteasomal degradation, as verified *in vivo* via CHFR^{-/-} mice (Kashima *et al.*, 2012).

Activation of the ubiquitination enzymatic cascade at locations of DSBs also facilitates the localization of downstream DNA repair proteins (such as BRCC36 and Rap80) (Wang and Elledge, 2007). Correlation in RNF8 and CHFR-specific E2 ubiquitin conjugating enzymes responsible for DSB-specific E3 recruitment and initiation of the ubiquitination cascade suggest a shared function by both E3 ubiquitin ligases in mediating the first ubiquitination-mediated DNA damage response in cells (Liu *et al.*, 2013, Wang and Elledge, 2007).

1.3.3 CHFR down-regulation via promoter hypermethylation and mutation in cancer cell lines and tumours

Down regulation of CHFR expression via promoter hypemethylation is frequently identified in numerous cancer cell lines and primary tumours; including well-characterised cancer cell lines, such as DLD1, HCT116 and IMRS (Scolnick and Halazonetis, 2000) (Table 1.2).

Table 1.2 Selected summary of *CHFR* promoter hypermethylation and down regulation identified in primary cancers and cell lines.

	Percentage of cell lines or tumours *	Citation
Lung (NSCLC)	19.0	Mizuno <i>et al.</i> , 2002
Oesophagus	26.0	Shibata <i>et al.</i> , 2002
Lung (NSCLC)	10.0	Corn <i>et al.</i> , 2003
Colon	80.0	
Brain	100.0	
Bone	100.0	
Colorectal	53.0	Toyota <i>et al.</i> , 2003
Head/ neck	30.0	
Gastric	39.0	Satoh <i>et al.</i> , 2003
Breast	50.0	Erson and Petty, 2004

* in comparison to normal cell or tumour counterpart controls. NSCLC: non small-cell lung cancer.

CHFR mutations within primary cancers and tumours have also been observed. Scolnick and Halazonetis (2000) identified a loss-of-function mutation (V580M) within the C-terminal cysteine rich domain of a U2OS (osteosarcoma) cell line within a CG-dinucleotide. In addition, three mutations (in non-small cell lung cancer patients) within the *chfr* gene were identified by Mariatos *et al* (2003) were previously identified, including 2 missense mutations and a loss of heterozygosity identified in the third. Loss of CHFR's function as a checkpoint protein in mitotically dividing cells therefore leads to mitotic progression from the G₂ to M transition (Erson and Petty, 2004, Privette *et al.*, 2008; Section 1.3.1), despite any adverse conditions within the cell (such as DNA DSB, See Section 1.3.2).

1.3.4 Domain architecture

Domain architecture of the CHFR protein consists of an N-terminal Forkhead Associated (FHA) domain, central RING domain and a C-terminal cysteine-rich

domain, with numerous homologues and paralogs identified within a range of different organisms (Brooks *et al.*, 2008) (Figure 1.8).

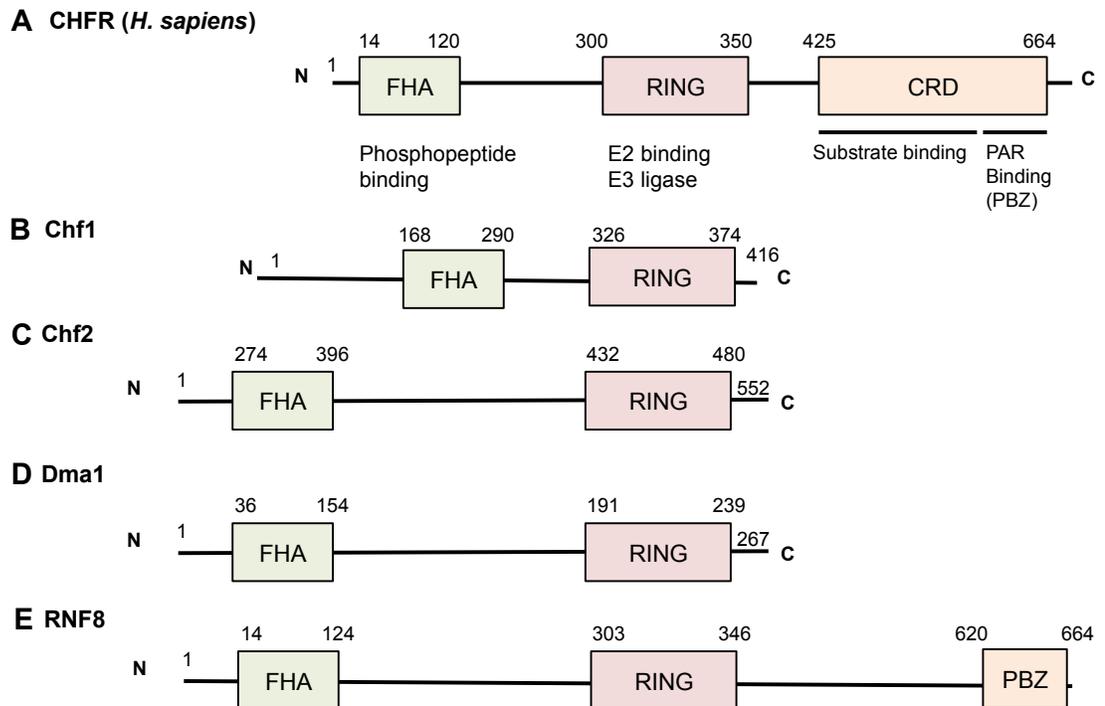


Figure 1.8 Domain architecture of the full-length CHFR protein, in comparison to protein paralogs and homologs (adapted from Brooks *et al.*, 2008 and Oberoi *et al.*, 2010). (A) Domain architecture and roles of individual CHFR domains are indicated (panel adapted from Oberoi *et al.*, 2010). Residue numbers for the human full-length CHFR (FL-CHFR) protein are shown corresponding to isoform 2 (Accession number: NP_001154817.1); whereby an N-terminal FHA, central RING and C-terminal cysteine rich domain (CRD), including a Poly-(ADP)ribose (PBZ) binding domain, are indicated (Oberoi *et al.*, 2010). Pairwise alignments of CHFR homologues and paralogs conducted by Brooks *et al.* (2008) identified that (B) Chf1 and (C) Chf2 *S. cerevisiae* paralogs also contain an N-terminal FHA and central RING domain. (D) Dma1 of *S. pombe* (Murone and Simanis, 1996) and (E) human RNF8 share an identical conservation of FHA and RING domain conservation to (A) CHFR, (B) Chf1 and (C) Chf2 (Brooks *et al.*, 2008). The human RNF8 homologue of the CHFR protein also contains (E) a PBZ domain (Brooks *et al.*, 2008).

FHA and RING domain-containing homologues frequently exhibit activities mediating cell cycle regulation (Fukuda *et al.*, 2008). For example, the FHA-RING E3 ligase Defective mitotic arrest 1 (Dma1) of *Schizosaccharomyces pombe* (*S. pombe*) is responsible for the inhibition of cytokinesis in response to incorrect chromosomal attachment to the mitotic spindles (Murone and Simanis, 1996; Johnson and Gould, 2011). Proceeding (obligate) Dma1 homodimerization

(Johnson *et al.*, 2012) and ubiquitination of the Sind4 scaffold protein, translocation of the Separation Initiation Network (SIN; Guertin *et al.*, 2002; Johnson and Gould, 2011) protein complex and Plo1 mitotic kinase to the microtubules is inhibited, thus preventing progression of cytokinesis (Johnson and Gould, 2011).

Paralogs of the *S. pombe* Dma1 FHA-RING E3 ubiquitin ligase, Dma1 and Dma2 (from *Saccharomyces cerevisiae*) are responsible for the ubiquitination, proteasomal degradation and subsequent down regulation of septins, responsible for GTP-dependent formation of filaments (Estey *et al.*, 2000; Chahwan *et al.*, 2013), spindle position and growth (Chahwan *et al.*, 2013; Versele and Thurner, 2005).

Whilst experimental structures have revealed obligate dimerization underpins the E3 ubiquitin ligase activities of both Dma1 (Johnson *et al.*, 2012) and RNF8 (Campbell *et al.*, 2012), the oligomeric state of the FL-CHFR protein and structural mechanisms underpinning potential multimer formation remains unknown. In addition, implications of oligomerization upon its E3 ubiquitin ligase activities remain unexplored.

1.3.4.1 The C-terminal cysteine rich domain (CRD)

The CHFR C-terminal cysteine rich domain (CRD) features both regions associated with substrate binding and a Poly(ADP)-ribose (PAR) specific region of interaction (Figure 1.8) (Oberoi *et al.*, 2010). A product of Poly(ADP-ribose) polymerase-mediated ADP-ribosylation, the PAR polymer is via the modification of NADP and has significant implications in cell survival, apoptosis and numerous mitotic regulatory activities (Schrieber *et al.*, 2006).

An X-ray structure of the CHFR C-terminal CRD was previously obtained by Oberoi *et al.* (2010). Featuring 4 zinc-binding motifs and 5 co-ordinated zinc ions, the structure of the CHFR C-terminus is highly conserved between human CHFR

(Oberoi *et al.*, 2010) and the murine homologue (Scolnick and Halazonetis, 2000) (Figure 1.9).

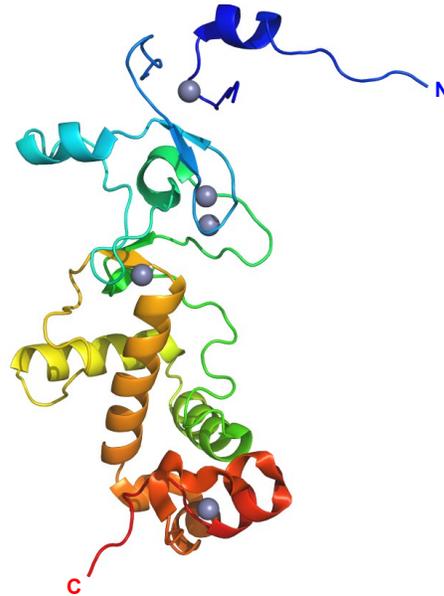


Figure 1.9 An X-ray structure of the human CHFR C-terminal cysteine rich domain (CRD) (Adapted from Oberoi *et al.*, 2010, PDB: 2XP0). Composed of 4 β -strands and 8 α -helices, the CHFR C-terminus contains 5 zinc ions (Zn^{2+}) co-ordinated to 4 different zinc ion binding centres (Oberoi *et al.*, 2010). Zinc ions are shown as grey spheres whilst the protein is coloured within a spectrum from navy to red, corresponding to the N to C-terminus (residues 425 to 663, respectively).

Interestingly, the CHFR C-terminus crystal structure was crystalized within a C2 (dimeric) space group (Oberoi *et al.*, 2010), with some crystal contacts identified. However, it remains unclear as to whether the solution structure of the protein is dimeric; with any implications concerning oligomeric state of the CHFR protein and its role in its E3 ubiquitin ligase activities remaining unclear.

Using deletion mutants, the CHFR: Aurora A interaction was specifically mapped to the CHFR CRD and the Aurora A N-terminus (Yu *et al.*, 2005), potentially targeting the mitotic kinase for proteasomal degradation.

With interactions mapped to the CRD (via the use of domain deletion mutants), CHFR is also responsible for the ubiquitination and targeted proteasomal degradation of the chromosomal remodelling complex: Helicase-Like Transcription Factor (HLTF), with HLTF autoubiquitination verified *in vitro* (Kim *et al.*, 2010). In

the presence of CHFR, destabilisation of HLTF also leads to the down-regulation of Plasminogen-Activation Inhibitor (*PAI-1*) expression (Kim *et al.*, 2010; Lauffenburger *et al.*, 1996; Keskiöja *et al.*, 1987), an important regulator of tumour invasion and metastasis (Lauffenburger *et al.*, 1996; Keskiöja); indicative of a tumour suppressive role for the CHFR protein (Kim *et al.*, 2010). A tumour suppressive role for the CHFR protein was identified by Oh *et al.* (2009). Through interacting with the CHFR CRD, Histone Deacetylase-1 (HDAC-1) is polyubiquitinated (*in vivo* and *in vitro*) and targeted for proteasomal degradation, with tumour suppression and reduction in metastasis also observed.

1.3.4.2 The N-terminal FHA domain

Evolutionary conserved within a wide range of organisms (Hoffmann and Bucher, 1995; Durocher, *et al.*, 1999) and typically identified within protein transcription factors, FHA domains are associated with DNA binding (Ding *et al.*, 2005; Hoffmann and Bucher, 1995). Spanning 65 to 75 residues (Durocher, *et al.*, 1999), typical structures include numerous antiparallel β -strands and a centrally conserved histidine residue (also known as a β - sandwich fold) and three discrete motifs of conserved residues, separated by a central, highly variable region (Hoffmann and Bucher, 1995). NMR relaxation experiments by Ding *et al.* (2005) have identified significant rigidity in regions of the *Arabidopsis*-kinase associated protein phosphatase (AKPP) FHA domain, with fluctuations mapped to regions shared within the CHFR FHA domain crystal structure counterpart (Ding *et al.*, 2005; Stavridi *et al.*, 2002).

The CHFR FHA domain crystal structure was obtained by Stavridi *et al.* (2002) (Figure 1.10, Panel A).

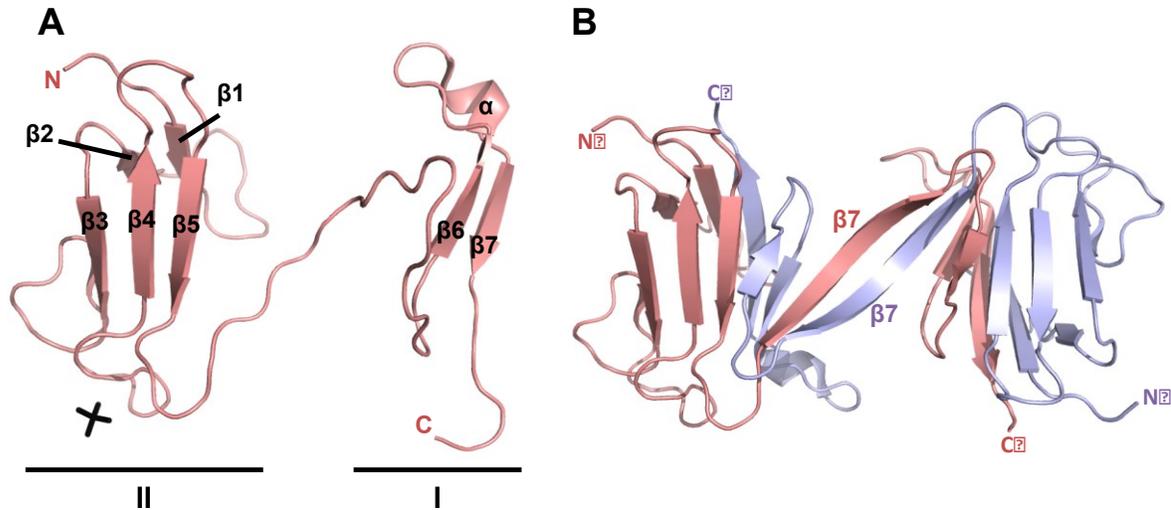


Figure 1.10 Monomeric and segment swapped dimeric experimental structures of the CHFR FHA domain (Adapted from Stavridi *et al.*, 2002). (A) The individual CHFR FHA domain monomer (PDB: 1LGP) can be split into two discrete segments. In the first (I), a short, central α -helix positioned between two β -strands located at the C-terminus, with 5 β -strands in the second (II) segment, located closer to the N-terminus (Stavridi *et al.*, 2002). Tungstate, a charge and size comparable analogue of a phosphate (Stavridi *et al.*, 2002), is indicated by the stick. (B) Within the FHA domain segment swapped dimer (PDB: 1LGQ), 2 regions of each domain monomer contributes half the N-terminus of β -strand 6 and half the N-terminus of β -strand 7 to the domain of the other (Stavridi *et al.*, 2002). Individual domains are indicated in salmon and blue, respectively. For clarity, only β -strand 7 per FHA monomer is indicated.

Interestingly, a β - sandwich fold, similar to that identified within the AKPP FHA domain homologue (Ding *et al.*, 2005), is evident within the CHFR FHA domain; with segment swapping and dimerization also exhibited both within the X-ray crystal structure (Figure 1.10, Panel B) and Analytical Ultra Centrifugation (AUC) data (Stavridi *et al.*, 2002). However, Stavridi *et al.* (2002) were unable to specify (via both AUC and X-ray crystallography experiments) whether the segment swapped dimerization observed was an artefact of the crystallization process via superimposition of individual domains, or a viable dimeric conformation for the CHFR RING domain in solution. The oligomerization state of the FL-CHFR protein currently remains unknown. In addition, potential dimerization of the FL-CHFR protein via the FHA domain could hold implications, with regards to CHFR ubiquitination activities and substrate recognition.

Co-crystallization of the CHFR N-terminal FHA domain with tungstate further supports the postulation that CHFR can directly interact with phosphopeptides (Stavridi *et al.*, 2012), potentially mitotic kinases (such as Aurora A and PLK1; Kang *et al.*, 2002; Yu *et al.*, 2005) implicated in cell cycle regulation via CHFR-mediated ubiquitination (Figure 1.7), but this has yet to be fully explored, in both structural and intracellular studies. However, using FHA domain CHFR deletion mutants within *in vitro* assays, Fukuda *et al.* (2008) have mapped CHFR's anti-proliferative specifically to the N-terminal FHA domain, suggesting some cell cycle mediated control is implicated specifically to the protein's N-terminus.

1.3.4.3 Central RING domain

The CHFR central RING domain (Figure 1.8, Panel A) contains the typical cross-brace motif (Scolnick and Halazonetis, 2000), exhibited within other RING E3 ubiquitin ligase proteins (Figure 1.8), catalysing the direct transfer of ubiquitin within the E2~ubiquitin conjugate to a target protein for ubiquitination (Kang *et al.*, 2002, Chaturvedi *et al.*, 2002), or to one of its own ϵ -amino groups via autoubiquitination (Figure 1.2, Bothos *et al.*, 2003). As of September 2017, no experimental structure or model of the central CHFR RING domain is available.

Numerous *in vitro* and *in vivo* ubiquitination assays with CHFR RING deletion mutants (Chaturvedi *et al.*, 2002; Kang *et al.*, 2002; Bothos *et al.*, 2003) have verified that the central CHFR RING domain alone (spanning a minimum of residues 267 to 360; Kang *et al.*, 2002) is essential in facilitating its E3 ubiquitin ligase activities. Therefore, the C-terminal CRD and N-terminal FHA domain are not required for CHFR's autoubiquitination *in vitro* (Kang *et al.*, 2002).

However, the CHFR FHA domain may have a regulatory role concerning its *in vitro* autoubiquitination activities (Kang *et al.*, 2002). Kinetic studies performed to assess CHFR's autoubiquitination rates in a CRD deletion mutant with an FHA domain (Δ CRD-CHFR, residues 1 to 360) compared to two FHA- CRD deletion mutant

proteins (Δ FHA- Δ CRD-CHFR) identified a higher rate of autoubiquitination in assays containing Δ FHA- Δ CRD-CHFR E3 ligases (spanning residues 142 to 360 or 267 to 360) compared to the Δ CRD-CHFR counterpart (Kang *et al.*, 2002). In addition, the work of Scolnick and Halazonetis (2002) has identified a dominant-negative effect of transfected Δ FHA-CHFR proteins upon endogenous, FL-CHFR within a DLD1 cell line. Whilst Kang *et al.* (2002) identified a higher rate of CHFR autoubiquitination within Δ FHA- Δ CRD-CHFR mutants *in vitro*, *in vivo* assays by Scolnick and Halazonetis (2002) identified a dominant-negative role for FHA domain deletions mutants. A SAOS2 cell line transfected with a Δ FHA-CHFR plasmid (2-142), endogenous and FL-CHFR lost its capacity to delay mitotic progression at the antephasic checkpoint; whereby transfected cell lines exhibited up to a 5 times higher mitotic index, in comparison to non-transfected (FL-*chfr* expressing), transfected (with FL-CHFR) and vector controls (Scolnick and Halazonetis, 2002), with similar loss of the antephasic checkpoint observed in Ptk1 cells transfected with the dominant negative Δ FHA-CHFR protein (Matsusaka and Pines, 2004).

Functional E2 ubiquitin conjugating enzymes responsible for CHFR include: Ubc4, Ubc5A (Chaturvedi *et al.*, 2002; Kang *et al.*, 2002) and Ubc5B (Kang *et al.*, 2002; Bothos *et al.*, 2003), the Ubc13: Mms2 heterodimer (Bothos *et al.*, 2003). More specifically, tandem mass spectrometry (LC/MS/MS) has revealed CHFR is autoubiquitinated in the presence of a UbcH5 isoform (unspecified); with K11 linkages associated with the production of high-molecular weight polyubiquitin chain products; in conjunction with a low molecular weight sub-population of K48 and K63 conjugates (Jung *et al.*, 2013). With such a limited number of E2 ubiquitin-conjugating enzymes confirmed as substrates for CHFR autoubiquitination, knowledge concerning CHFR's preference for E2s in initiating its E3 ubiquitin ligase activity is severely limited and any evolutionary conservation concerning active E3s remains unexplored.

Ubiquitination assays inclusive of the UbcH10 (Kang *et al.*, 2002; Chaturvedi *et al.*, 2002), Ubc2 (Chaturvedi *et al.*, 2002) and UbcH7 (Kang *et al.*, 2007; Bothos *et al.*, 2003) and UbcH8 (Bothos *et al.*, 2008) resulted in no formation of high molecular weight polyubiquitin chains (Kang *et al.*, 2002). In addition, individual E2s Ubc13, Mms2 do not result in CHFR autoubiquitination, whereas the Ubc13: Mms2 E2 heterodimer is required for E3 ubiquitin ligase activities (Bothos *et al.*, 2003), corroborating with other RING E3 ubiquitin ligase activities (Campbell *et al.*, 2012, Branigan *et al.*, 2012). By titrating increasing concentrations of E3 present and observing the amount of polyubiquitin chain product formed, CHFR has a notable preference for the Ubc13: Mms2 heterodimer, in comparison to the UbcH5B E2 enzyme (Bothos *et al.*, 2003).

By assessing the residue conservation between CHFR, APC11 and C-Cbl RING E3 ubiquitin ligases, Kang *et al* (2002) introduced two mutations (I306A and W332A) to ablate CHFR autoubiquitination *in vitro* and *in vivo* by performing a multiple sequence alignment between both E3 RING homologues and predicting evolutionary conserved (and therefore, potentially structurally and biochemically important) residues (Scolnick and Halazonetis, 2000).

However, since no experimental structure or model of the CHFR RING domain is currently available, the molecular basis of interactions underpinning all interfaces within the CHFR RING: E2~ ubiquitin complex (particularly CHFR RING: E2 and CHFR RING: ubiquitin interfaces) have yet to be modelled. In addition, biochemical verification of all interface interactions within the CHFR ubiquitination complex has yet to be performed, particularly concerning modelling such interactions within a structural and biochemical, rather than evolutionary, context.

By coupling immunoaffinity assays and mass spectrometry, Oh *et al* (2007) identified USP7 as a DUB enzyme responsible for stabilizing the CHFR protein and cleaving polyubiquitin chains from the autoubiquitinated E3 ligase; with *in vivo* accumulation of CHFR attributed to lack of protein translocation for proteasomal

degradation. In addition, Kwon *et al* (2012) propose a UBC9-mediated model for the posttranslational modification of CHFR via the attachment of the protein SUMO (via SUMOylation), potentially priming CHFR for ubiquitination via an independent E3 ligase. Subsequent SUMOylation and ubiquitination of CHFR may contribute to its targeted proteasomal degradation (Kwon *et al.*, 2012).

1.4 AIMS AND OBJECTIVES

Within adverse conditions, CHFR delays mitotic progression at the antepause checkpoint by the ubiquitination of substrates. In addition, CHFR itself can be subjected to autoubiquitination. While a small number of E2 ubiquitin conjugating enzymes responsible for ubiquitination of substrate proteins and/or CHFR autoubiquitination have been identified, the basic biophysical characteristics of the CHFR protein, structure of central RING domain and the biochemical basis of CHFR RING: E2 ~ubiquitin complex formation remains very poorly understood.

Within in the present study, the following aims and objectives were set:

1. Determine the oligomeric state of the FL-CHFR protein in solution and investigate the structural features responsible for any multimer formation:

- Recombinantly express CHFR domain deletion mutant proteins, with molecular weights determined using protein standards (with SEC) and native PAGE.
- Predict coiled-coil content within the FL-CHFR protein using servers.
- Verify the molecular weight, oligomeric state, shape and size of the FHA domain (in solution) using SEC-SAXS.
- Explore the ubiquitination activities of verified oligomers and domain deletion mutants.

2. Identify new E2 enzymes responsible for CHFR- mediated polyubiquitin chain formation *in vitro*:

- Perform ubiquitination assays using multiple E2 enzymes.

- Explore evolutionary relationship between E2s via phylogenetic analysis.
3. Model the structure of the central RING domain:
- Use the experimental structures of E3 RING homologues to produce a homology model.
 - *In silico*, explore the dynamics of the protein model's backbone and individual residues using Molecular Dynamic (MD) simulations.
4. Model and verify key interactions within the CHFR RING: Ubc13~ ubiquitin complex:
- Perform detailed bioinformatics analysis of E3: Ubc13 (: Mms2) ~ ubiquitin and E3: Ubc13 ~ ubiquitin experimental structures to model CHFR RING: Ubc13 and CHFR RING: ubiquitin interactions.
 - Recombinantly produce FL-CHFR (RING domain) and ubiquitin site directed mutants to target interactions.
 - Verify disruption of targeted interactions via ubiquitination assays using site directed FL-CHFR (RING) and ubiquitin proteins.

CHAPTER 2: Biochemical and biophysical characterisation of CHFR

CHAPTER 2 INTRODUCTION

The structure of the FL-CHFR protein consists of an evolutionary conserved N-terminal FHA domain, central RING domain and C-terminal cysteine rich (CRD) domain (Scolnick and Halazonetis, 2002) (Chapter 1, Figure 1.8). Molecular organization, with regards to the CHFR FHA and RING domains, is comparable to two different yeast orthologs, including the obligate dimer Dma1 (Chapter 1, Figure 1.8; Johnson *et al.*, 2010).

Whilst segment swapped dimerization of the N-terminal FHA domain has been observed within both a crystal structure and Analytical Ultra Centrifugation (AUC) gradient (inclusive of monomeric and dimeric oligomeric subspecies (Chapter 1, Figure 1.10)), conformation of FHA domain dimer in solution and verification of segment swapping (Stavridi *et al.*, 2002) has yet to be performed within solution using an additional independent method. In addition, crystal contacts observed in the C2 space group modelled via the CHFR C-terminal CRD crystal structure (Chapter 1, Figure 1.9) (Oberoi *et al.*, 2010) suggest the C-terminus may be dimeric in solution. An understanding of the FL-CHFR oligomeric state in solution and that of the individual protein domains (FHA and C-terminus), including domain deletion mutants, may provide important information regarding CHFR's oligomeric state in solution and a potential obligatory role in mediating its formation of polyubiquitin chains *in vitro*.

The following chapter describes experiments exploring the oligomeric state of numerous CHFR protein constructs in solution, including calculations of CHFR protein molecular weights and native PAGE analysis of proteins. Coiled-coil prediction tools were used to identify potential structural features responsible for higher order oligomeric states resultant of heptad-repeats. In addition, a low-resolution model of the FHA domain in solution was used to compare the protein dimensions of the FHA domain within the X-ray crystallography data set (Stavridi *et*

al., 2002) against the solution counterpart. Potential domain-specific roles for CHFR-mediated polyubiquitin chain formation have also been explored.

In addition, a severely limited number of E2 ubiquitin conjugating enzymes have been previously identified as responsible for CHFR mediated polyubiquitin chain formation *in vitro* (Chapter 1, Section 1.3.4.3; Chaturvedi *et al.*, 2002; Kang *et al.*, 2002; Bothos *et al.*, 2003, Jung *et al.*, 2013). To explore a wider range of E2s responsible for polyubiquitin chain formation and important implications concerning both polyubiquitin chain linkage types (Rape and Ye, 2009) and downstream verified functions (Section 1.2.3, Table 1.1), a selection of E2 enzymes have been screened to further characterize CHFR's specificity for selected E2s. Phylogenetic information has also provided an evolutionary context to E2 sequence similarity and respective capacities of CHFR to mediate polyubiquitin chain formation *in vitro*.

2.1 MATERIALS AND METHODS

A detailed description of chemicals, biochemicals and vectors is provided in Appendix 1, Section A1.1.

2.1.1 Antibodies

Primary antibody against ubiquitin (mouse monoclonal, IgG1, ab7254, Abcam) and a secondary anti-mouse antibody (ECL anti-mouse IgG horseradish peroxidase linked antibody from sheep, GE Healthcare) were diluted 1: 500 and 1: 2000 (respectively) into a blocking buffer (5 % w/v skimmed milk powder, 1X PBS and 0.1 % v/v Tween 20).

A primary antibody against CHFR was also used (rabbit polyclonal, IgG, PA5-28079, ThermoScientific), raised against a synthetic peptide corresponding to the CHFR C-terminus (residues 602-664). It was diluted 1: 2000 into a blocking buffer (5 % w/v skimmed milk powder, 1X PBS and 0.1 % v/v Tween 20) and used with a

second anti-rabbit antibody (ECL anti-rabbit IgG horseradish peroxidase linked antibody from donkey, GE Healthcare); diluted (1: 2000) in an identical blocking buffer.

2.1.2 Recombinant protein expression and purification

2.1.2.1 Expression of FL-CHFR, CHFR Δ FHA 1 to 3, FHA and CHFR C-terminus proteins

20 μ l BL21 CodonPlus (DE3) RIL cells (Appendix 1, Section A1.2) were thawed on ice, with 1 μ l (150-200 ng) bacterial expression vector (Appendix 1, Table A1.2) and incubated on ice for a further 30 minutes. Cells were then heat-shocked at 42 °C for 42 seconds, followed by an additional 2 minute incubation on ice. 200 μ l LB was added and then transformations incubated (37 °C, 5.844 x g; Infors HT, Ecotron) for an hour. 200 μ l of each transformation was then plated onto room-temperature LB-agar/ antibiotic plates (Table 2.1) and incubated overnight at 37 °C.

Table 2.1 LB-agar antibiotic concentrations for bacterial selection.

Antibiotic	[Stock], mg/mL	[working], mg/mL
Kanamycin sulphate	50 mg/mL	50 μ g/mL
Chloramphenicol	34 mg/mL (dissolved in ethanol)	34 μ g/mL

A pre-culture consisted of 100 mL of LB media supplemented with antibiotics kanamycin and chloramphenicol, with final concentrations of 50 and 34 μ g/mL, respectively; inoculated a single colony and incubated overnight (37 °C, 5.844 x g; Infors HT, Ecotron).

Within a 2L conical flask, 1L of LB (supplemented with correct antibiotics, Table 2.1), 10mL of pre-culture was added and incubated (37 °C, 5.844 x g; Infors HT,

Multitron) to an optical density (at 600 nm, OD₆₀₀) of 0.4 (for FL-CHFR, ΔFHA and C-terminus CHFR proteins) or 0.6 (for expression of CHFR FHA domain).

0.4 mM zinc chloride was added to FL-CHFR, ΔFHA and C-terminus CHFR protein cultures just prior to induction with 0.4 mM IPTG (or 0.5 mM for FHA CHFR), followed by an overnight incubation at 21 °C (5.844 x g).

Cultures were then centrifuged within a JLA 8.1000 fixed angle rotor (Beckman Coulter) at 5,044.081 x g (4,500 RPM at r_{max} of 222.8 mm) in an Avanti J26-XP centrifuge (Beckman Coulter) for 15 minutes (4 °C). Supernatant was then discarded and pellets briefly centrifuged (5 minutes at 4°C) within a SX4750 swinging bucket rotor at 2,890 x g (3,750 RPM at r_{max} of 207.8 mm; Allegra X-12R centrifuge, Beckman Coulter) and then stored at -80°C.

2.1.2.2 Purification of FL-CHFR

2.1.2.2.1 Ion exchange chromatography of FL-CHFR

FL-CHFR pellets were thawed on ice and resuspended in 200 mM NaCl, 50 mM Tris-HCl (pH 8.0), 5 % (v/v) glycerol and 10 mM BME; with a single (complete) protease inhibitor cocktail tablet added (EDTA-free, Sigma). Resuspended pellets were then sonicated on ice for 2 minutes (10 seconds on, 10 seconds off) with 40 % power (Misonix Sonicator 3000, Cole-Parmer). Clarification of cell lysates then followed via centrifugation using a JA 25.50 fixed angle rotor at 43,588 x g (19,000 RPM at r_{max} of 108 mm) within an Avanti J-20 XP centrifuge (Beckman Coulter) for 45 minutes at 4 °C.

The retained supernatant was then filtered (0.45 μm Minisart, Sartorius) and manually injected onto a washed (in distilled water) and pre-equilibrated (in lysis buffer) HiTrap Heparin HP column (5mL, GE Healthcare). The immobilized heparin functions as a high capacity cation exchanger via the anionic sulfate groups

present. The initial purification of clarified lysates (containing the strep II-tagged FL-CHFR) using the HiTrap Heparin HP column significantly improved protein purity by eliminating low molecular weight contaminants, in comparison to purification of lysates by immediate affinity chromatography using a Strep-*tactin* column. Prior to first usage, the column was washed with 25 mL MilliQ water by manual injection.

An ÄKTA Prime Plus chromatography system (GE Healthcare) was washed with the original filtered and de-gassed lysis buffer (Pump and Buffer A, 200 mM NaCl, 50 mM Tris-HCl (pH 8.0), 5 % (v/v) glycerol and 10 mM BME) and a high-salt elution buffer (Pump and Buffer B, 1 M NaCl, 50 mM Tris-HCl (pH 8.0), 5 % (v/v) glycerol and 10 mM BME). The HiTrap Heparin column (containing cell lysate) was then attached to the system and impurities washed off with Buffer A until the absorbance at 280 (A_{280}) was 0 mAu. Proteins were then eluted with a 80-100 mL gradient of 200 mM to 1 M NaCl (at 3 mL/ min) via incrementally increasing the amount of buffer B added to buffer A (Appendix 1, Section A1.4, Figure A1.1, Panel A). At approximately 500 mAu, the UV peak corresponds to the FL-CHFR protein, with numerous impurities (via non-specifically bound proteins) still remaining and as determined by SDS-PAGE (Appendix 1, Section A1.4, Panel B).

The HiTrap Heparin column was then cleaned (manually via syringe injection) in 25 mL MilliQ water, 50 mL 20 % (v/v) ethanol and stored in 20 % (v/v) ethanol at 4°C.

Fractions across the UV peak, corresponding to the HiTrap Heparin elution, were then analyzed by SDS-PAGE (Section 2.1.2.3) and all fractions within the elution pooled.

2.1.2.2.2 Affinity purification of FL-CHFR

Pooled fractions from ion exchange chromatography using the HiTrap Heparin column (Appendix 1, Section A1.4, Figure A1.2, Lane 1), containing the FL-CHFR

protein and other non-specific proteins, was then purified by affinity chromatography (via manual injection) using a 5 mL Strep-*tactin* Superflow Plus cartridge (Qiagen). Prior to first use, the column was washed with 25 mL MilliQ water, followed by equilibration with the original lysis buffer (A; 200 mM NaCl, 50 mM Tris-HCl (pH 8.0), 5 % (v/v) glycerol and 10 mM BME). The pooled HiTrap Heparin elution was then passed 4 times through the column and unbound proteins removed via four washes with buffer A (50 mL); each retained for analysis by SDS-PAGE (Appendix 1, Section A1.4, Figure A1.2, Lanes 2 to 5).

The FL-CHFR protein was then eluted with 50 mL buffer B (elution buffer; 200 mM NaCl, 50 mM Tris-HCl (pH 8.0), 5 % (v/v) glycerol, 10 mM BME and 3mM d-Desthiobiotin) and analyzed by SDS-PAGE (Section 2.1.2.3; Appendix 1, Section A1.4, Figure A1.2, Lane 6).

The Strep-*tactin* Superflow cartridge was regenerated by manually injecting the column with the following (in chronological order): 25 mL MilliQ water, 25 mL 1mM 2-(4-Hydroxyphenylazo)-benzoic acid (HABA), 25 mL MilliQ water, 1 M NaOH and 25 mL MilliQ water. The column was then stored in 20 % (v/v) ethanol at 4 °C.

2.1.2.2.3 Size Exclusion Chromatography (SEC) of FL-CHFR

The Strep-*tactin* Superflow elution (containing the FL-CHFR protein, Appendix 1, Section A1.4, Figure A1.2, Lane 6) was then concentrated to 5 mL using a Vivaspin Protein Concentrator spin column (30,000 MWCO, Sigma); with spin columns washed with 30 mL MilliQ water, prior to first use. Spin columns were centrifuged using a SX4750 swinging bucket rotor at 2,890 x g (3,750 RPM at r_{\max} of 207.8 mm) within a Allegra X-12R centrifuge (Beckman Coulter) at 4 °C.

A Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare) was equilibrated with 280 mL of gel filtration (GF) buffer (150 mM NaCl, 25 mM Tris-HCl (pH 8.5), 2 % (v/v) glycerol) using an ÄKTA Prime Plus chromatography system

(GE Healthcare). A 5 mL sample injection loop (GE Healthcare) was filled (by manual injection) with 10 mL GF buffer, followed by the 5mL FL-CHFR protein sample.

The protein sample was then injected (via ÄKTA Prime Plus) onto the column at 0.5 mL/ minute (for 15 mL), followed by loading of ~195 mL GF buffer (Section A1.4, Figure A1.3, Panel A). Selected fractions were analyzed by SDS-PAGE (Section 2.1.2.3; Appendix 1, Section A1.4, Figure A.13, Panel B), pooled and protein concentration determined via The Beer-Lambert Law, substituting a path length of 1.0mm (not shown):

$$C = A / \epsilon$$

Where C is concentration (in molar, M), A is absorbance at 280 nm and ϵ is the wavelength-dependent molar absorptivity coefficient (or extinction coefficient) of Strep(II)-tagged full-length CHFR ($72310 \text{ M}^{-1} \text{ cm}^{-1}$). The extinction coefficient and estimated molecular weight of all proteins was calculated by entering the full amino acid sequence within the ProtParam Server (Version 1.1; Gasteiger *et al.*, 2005).

The strep(II)-tagged full length (1-664) CHFR protein was then flash-frozen in liquid nitrogen and stored at -80°C .

2.1.2.3 SDS-PAGE

Protein samples were resolved using Novex NuPAGE, 4-12 % BisTris (polyacrylamide) gels (in 1X MES running buffer, both Life Technologies) with 10 μL loading dye (50 mM Tris-HCl pH 7.5, 2 % v/v SDS, 10 % v/v glycerol, 2 % v/v BME, 12.5 mM EDTA and 0.02 % v/v bromophenol blue) added to a 20 μL fraction sample and heated at 98°C for 5 minutes prior to loading. A molecular weight marker (3 to 198 kDa; SeeBlue Plus2 Pre-Stained Standard, Thermo Fisher Scientific) was also included.

Gels were resolved within an XCellSureLock Mini-Cell (Life Technologies) at (a constant) 200V for 30 minutes and destained over 24 hours (InstantBlue™, Expedeon); with residual destaining reagent removed with MilliQ water, prior to scanning.

2.1.2.4 Purification of CHFR Δ FHA proteins

Purification of CHFR Δ FHA1 (231-664) (Appendix 1, Section A1.4, Figures A1.4 and A1.5), CHFR Δ FHA2 (251-664) (Appendix 1, Section A1.4, Figures A1.6 and A1.7) and CHFR Δ FHA3 (267-664) (Appendix 1, Section A1.4, Figures A1.8 and A1.9) proteins by ion exchange (Section 2.1.2.2.1), affinity (Section 2.1.2.2.2) and size exclusion (Section 2.1.2.2.3) chromatography was performed as previously described for FL-CHFR, with ion exchange chromatography performed by manual injection. Protein purifications were also resolved by SDS-PAGE analysis (Section 2.1.2.3) as previously described.

Proceeding size exclusion chromatography (Section 2.1.2.2.3), selected protein fractions were pooled and concentrated as previously described, with protein purifications resolved by SDS-PAGE analysis (Section 2.1.2.3). Protein concentrations were calculated using an extension coefficient for each CHFR Δ FHA protein and corresponding to the individual amino acid sequence.

The strep (II)-tagged CHFR FHA deletion mutant proteins (Δ FHA1, Δ FHA2 and Δ FHA3) were then flash-frozen in liquid nitrogen and stored at -80°C.

2.1.2.5 Purification of the CHFR FHA (13-180) domain

2.1.2.5.1 *Immobilized metal ion affinity chromatography of the CHFR FHA domain*

A CHFR FHA pellet was thawed on ice and resuspended in 250mM NaCl and 5mM DTT; with a single (complete) protease inhibitor cocktail tablet added (EDTA-free, Sigma). Resuspended pellets were then sonicated on ice for 2 minutes (10 seconds on, 10 seconds off) with 40 % power (Misonix Sonicator 3000, Cole-Parmer). Clarification of cell lysates was then performed by centrifugation using a JA 25.50 fixed angle rotor at 43,588.584 x g (19,000 RPM at r_{\max} of 108 mm) within an Avanti J-20 XP centrifuge (Beckman Coulter) for 45 minutes at 4 °C.

The retained supernatant was then filtered (0.45 μ m Minisart, Sartorius) and adjusted to 20mM imidazole. Prior to first usage, a 5 mL HisTrap FF column (GE Healthcare) was washed (in 25mL MilliQ) and pre-equilibrated in 25 mL wash buffer (A); consisting of 50 mM NaPO₄-HCl (pH 8.0), 150 mM NaCl and 20mM imidazole. The clarified lysate was then manually injected onto the column, proceeded by a 20 column volume (100 mL) wash to remove unbound, non-specific proteins.

N-terminal His-tagged FHA protein was then eluted by manual injection using 65 mL of a high-concentration imidazole elution buffer (B; 50 mM NaPO₄-HCl (pH 8.0), 150 mM NaCl and 250mM imidazole); with individual 5 mL fractions collected (Appendix 1, Figure A1.10, Panel A), resolved by SDS-PAGE (Section 2.1.2.3,) and selected fractions (Lanes 1-9) pooled for overnight TEV cleavage of the His-tag and dialysis.

The HisTrap FF column was stripped and recharged between all protein metal ion affinity chromatography-specific purifications as follows: 5 column volumes (CV, 25 mL) MilliQ water, 5 CV 1 M sodium hydroxide, 5 CV MilliQ water, 5 CV stripping buffer (20 mM NaPO₄, 50 mM NaCl, 50mM EDTA-NaOH pH 8.0), 5 CV MilliQ Water, 5 CV 0.1 M nickel (II) chloride and 5 CV MilliQ water; storing in 20 % (v/v) ethanol at 4°C.

2.1.2.5.2 His-tag cleavage, dialysis and second immobilized metal ion affinity chromatography of the CHFR FHA domain

Pooled CHFR FHA domain elutions from the initial HisTrap metal ion affinity chromatography elution (Section 2.1.2.5.1) were then subject to TEV-cleavage of the N-terminal His-tag, whereby His6-TEV (in 5% v/v glycerol) was added to a final concentration of 20 µg/mL.

The protein (Appendix 1, Section A1.4, Figure A1.10, Panel B, Lane 1) was dialyzed overnight within 10,000 MWCO SnakeSkin dialysis tubing (briefly hydrated in MilliQ water prior to filling, ThermoFisher Scientific) within a dialysis buffer (50 mM Tris-HCl pH 7.5, 200mM NaCl, 5mM MgCl₂, 5mM BME, 10% (v/v) glycerol) at 4°C for 18 hours; covered, on a magnetic stirrer and containing a magnetic ‘flea’ for continuous mixing. Cleavage of the N-terminal His-tag from the CHFR FHA protein via TEV protease recognition of the cleavage site (Asn-Leu-Tyr-Phe-Gln↓Gly) resulted in conservation of an additional glycine residue at the N-terminus of the FHA protein.

Proceeding cleavage of the His-tag from the CHFR FHA domain (Appendix 1, Figure A1.10, Panel B, Lane 2), the untagged FHA protein was then isolated from both the His-tagged TEV protease and uncleaved His-tagged FHA protein by a second metal ion affinity chromatography purification step. The dialyzed protein was manually injected onto a stripped, re-charged and re-equalibrated (in buffer A) HisTrap FF column, with the flow-through, containing the untagged CHFR FHA domain protein, retained in 5mL fractions (Appendix 1, Section A1.4, Figure A1.10, Panel B, Lanes 1 to 12). Fractions 1 to 9 were then retained for concentration and additional purification by size exclusion chromatography.

2.1.2.5.3 Size exclusion chromatography of the (untagged) CHFR FHA domain

The HisTrap FF flow-through (containing the untagged CHFR FHA protein) was then concentrated to 5 mL using a Vivaspin Protein Concentrator spin column (10,000 MWCO, Sigma); with spin columns washed with 30 mL MilliQ water, prior to first usage. Spin columns were centrifuged using a SX4750 swinging bucket rotor at $2,890 \times g$ (3,750 RPM at r_{\max} of 207.8 mm) within a Allegra X-12R centrifuge (Beckman Coulter) at 4 °C.

A Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare) was equilibrated with 280 mL of gel filtration (GF) buffer (150 mM NaCl, 25 mM Tris-HCl (pH 7.0), 2 % (v/v) glycerol) using an ÄKTA Prime Plus chromatography system (GE Healthcare). A 5 mL sample injection loop (GE Healthcare) was filled with 10 mL GF buffer (manually injected), followed by the 5mL FHA protein sample.

The protein sample was then injected (via ÄKTA Prime Plus) onto the column at 0.5 mL/ minute (for 15 mL), followed by loading of ~195 mL GF buffer (Appendix 1, Section A1.4, Figure A1.11, Panel A). Selected fractions were analyzed by SDS-PAGE (Section 2.1.2.3; Appendix 1, Section A1.4, Figure A1.11, Panel B), pooled and protein concentration determined as previously described (Section 2.1.2.2.3); substituting an extinction coefficient (corresponding to the untagged FHA protein) of $16960 \text{ M}^{-1} \text{ cm}^{-1}$.

The concentrated untagged CHFR FHA domain (18-130) protein was then flash frozen in liquid nitrogen and stored at -80°C prior to use.

2.1.2.6 Purification of the CHFR C-terminal cysteine rich domain (394-664)

2.1.2.6.1 Anion exchange chromatography purification of the CHFR C-terminus

Bacterial pellets containing the CHFR C-terminal protein were thawed on ice and resuspended in 200 mM NaCl, 5 mM Tris (pH 8.0), 5% (v/v) glycerol and 10 mM

BME; with a single (complete) protease inhibitor cocktail tablet added (EDTA-free, Sigma). Resuspended pellets were then sonicated on ice for 2 minutes (10 seconds on, 10 seconds off) with 40 % power (Misonix Sonicator 3000, Cole-Parmer). Clarification of cell lysates then followed via centrifugation using a JA 25.50 fixed angle rotor at $43,588.584 \times g$ (19,000 RPM at r_{\max} of 108 mm) within an Avanti J-20 XP centrifuge (Beckman Coulter) for 45 minutes at 4 °C.

The retained supernatant was then filtered (0.45 µm Minisart, Sartorius) and manually injected onto a washed (in distilled water) and pre-equilibrated (in lysis buffer) HiTrap ANX FF (High Sub) anionic exchange column. Prior to first use, the column was washed with 25 mL MilliQ water by manual injection.

An ÄKTA Prime Plus chromatography system (GE Healthcare) was washed into the original filtered and de-gassed lysis buffer (Pump and Buffer A, 200 mM NaCl, 5 mM Tris (pH 8.0), 5% (v/v) glycerol and 10 mM BME) and high-salt elution buffer (Pump and Buffer B, 1 M NaCl, 50 mM Tris-HCl (pH 8.0), 5 % (v/v) glycerol and 10 mM BME). The HiTrap ANX FF column (containing cell lysate) was then attached to the system and impurities washed off with Buffer A until the absorbance at 280 (A_{280}) was 0 mAu (at 3 mL/min). Proteins were then eluted with a 80-100 mL gradient of 200 mM to 1 M NaCl via incrementally increasing the amount of buffer B added to buffer A (Appendix 1, Section A1.4, Figure A1.12, Panel A). At approximately 600 mAu, the UV peak corresponds to the FL-CHFR protein, with numerous impurities (via non-specifically bound proteins) still remaining and as determined by SDS-PAGE (Appendix 1, Section A1.4, Figure A1.12, Panel B). All fractions within the elution pooled for His-NusA tag cleavage and overnight dialysis.

The HiTrap ANX FF column was then cleaned (manually via syringe injection) in 25 mL 1M NaCl, 25 mL MilliQ water, 25 mL 20 % (v/v) ethanol and stored in 20 % (v/v) ethanol at 4°C.

2.1.2.6.2 His-tag cleavage, dialysis and second anionic exchange chromatography of the CHFR C-terminus

The CHFR C-terminus elution from the initial HiTrap ANX FF anion exchange chromatography (Appendix 1, Section A1.4, Figure A1.12, Panel B) was then subject to TEV-cleavage of the N-terminal His-NusA tag. His6-TEV (in 5% v/v glycerol) was added to a final concentration of 20 µg/mL.

The protein was then dialyzed overnight within 10,000 MWCO SnakeSkin dialysis tubing (briefly hydrated in MilliQ water prior to filling, ThermoFisher Scientific) within a TEV-specific dialysis buffer (50 mM Tris-HCl pH 7.5, 200mM NaCl, 5mM MgCl₂, 5mM BME, 10% (v/v) glycerol) at 4°C for 18 hours; covered, on a magnetic stirrer and containing a magnetic 'flea' for continuous mixing (Appendix 1, Section A1.4, Figure A1.13, Panel B, Lane 1). Cleavage of the N-terminal His-NusA tag from the CHFR C-terminus protein via TEV protease recognition of the cleavage site (Asn-Leu-Tyr-Phe-Gln↓Gly) resulted in conservation of an additional glycine residue at the N-terminus of the purified protein.

Proceeding cleavage of the His-NusA tag from the CHFR C-terminus (Appendix 1, Section A1.4, Figure A1.13, Panel B, Lane 2), the untagged protein was then isolated from both the His-tagged TEV protease and uncleaved His-NusA tagged protein by a second anion exchange chromatography purification step (Appendix 1, Section A1.4, Figure A1.13, Panel A). The dialyzed protein was injected onto a cleaned and re-equalibrated (in buffer A) HiTrap ANX FF column using a 50 mL Super Loop (GE Healthcare), with the flow-through containing the untagged protein retained in 3mL fractions (Appendix Figure A2.13, Panels B, C and D) and pooled for additional purification via size exclusion chromatography. All purifications were resolved by SDS-PAGE (Section 2.1.2.3).

2.1.2.6.3 Size exclusion chromatography of the (untagged) CHFR C-terminus

The HiTrap ANX FF flow-through (containing the untagged CHFR C-terminus) (Appendix 1, Section A1.4, Figure A1.3) was then concentrated to 5 mL using a Vivaspin Protein Concentrator spin column (20,000 MWCO, Sigma); with spin columns washed with 30 mL MilliQ water, prior to first usage. Spin columns were centrifuged using a SX4750 swinging bucket rotor at $2,890 \times g$ (3,750 RPM at r_{\max} of 207.8 mm) within a Allegra X-12R centrifuge (Beckman Coulter) at 4 °C. A Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare) was equilibrated with 280 mL of gel filtration (GF) buffer (150 mM NaCl, 25 mM Tris-HCl (pH 7.0), 2 % (v/v) glycerol) using an ÄKTA Prime Plus chromatography system (GE Healthcare). A 5 mL sample injection loop (GE Healthcare) was filled with 10 mL GF buffer (manually injected), followed by the 5mL CHFR C-terminus protein sample.

The protein sample was then injected (via ÄKTA Prime Plus) onto the column at 0.5 mL/ minute (for 15 mL), followed by loading of ~195 mL GF buffer (Appendix 1, Section A1.4, Figure A1.14, Panel A). Selected fractions were analyzed by SDS-PAGE (Section 2.1.2.3; Appendix 1, Section A1.4, Figure A1.14, Panel B), pooled and protein concentration determined as previously described (Section 2.1.2.2.3); substituting an extinction coefficient (corresponding to the untagged CHFR C-terminus) of $42860 \text{ M}^{-1} \text{ cm}^{-1}$.

Concentrated proteins were flash frozen in liquid nitrogen and stored at -80°C prior to usage.

2.1.2.7 Western blots

Following selected assays or protein resolution by native (Section 2.1.3.2) or SDS (Section 2.1.2.3) PAGE; western blotting was also performed. A 'sandwich' consisting of blotting pads, Whatman paper, gel, nitrocellulose, Whatman paper and blotting pads (respectively) was assembled within an XCell II Blotting Module

(Life Technologies); with all components (7 x 9 cm) submerged in transfer buffer (Appendix 1, Section A1.2) prior to assembly.

Briefly, a blotting pad was inserted to the base of an XCell II Blotting Module (Life Technologies); proceeded by a single piece of Whatmann chromatography paper (3MM) and the SDS-PAGE or native PAGE gel. A piece of nitrocellulose membrane (Amersham Protran Premium NC Nitrocellulose blotting membrane, 0.20 μ M, GE Healthcare) was then placed on top, followed by another piece of chromatography paper and a blotting pad. The blotting module was filled with a sufficient number of blotting pads prior to closure, ensuring no gaps were present within the electrode. The central chamber of the module was then filled entirely with transfer buffer; whilst the outside was filled with MilliQ distilled water; transferring at 30 V for 1 hour.

Following transfer, the molecular marker used on Native PAGE transferred gels was immediately indicated by pencil, since the proteins present are unstained and are susceptible to being washed off throughout the blocking and washing procedures. This was not necessary for the SDS-PAGE (stained) molecular weight marker used. The transferred proteins upon the membrane were briefly visualized to check transfer quality using 20 mL Ponceau S solution (Sigma) for 5 minutes; proceeding by a brief wash in MilliQ water. Subject to satisfactory transfer, membranes were initially blocked via incubation with a blocking buffer (5 % w/v skimmed milk powder, 1X PBS and 0.1 % v/v Tween 20) with agitation for 30 minutes.

The primary antibody/ blocking solution against ubiquitin (1:500 dilution, 2 mL total volume; Section 2.2.1) was incubated with the (blocked) nitrocellulose membranes overnight (in a cold room) with agitation. The primary antibody/ blocking solution against CHFR (1: 2000 dilution, 2 mL total volume) was incubated with the blocked nitrocellulose membrane at room temperature (with agitation) for 1 hour. Excess primary antibody solution was then washed from the nitrocellulose membranes

using 100 mL washing buffer (1X PBS and 0.1 % v/v Tween 20) with agitation and at room temperature for 10 minutes. This was then removed and repeated twice more. The secondary (HRP-conjugated) antibody/ blocking solution against IgG rabbit or mouse primary antibodies (corresponding to nitrocellulose blots incubated with primary antibodies against CHFR or ubiquitin, respectively; 1: 2000 each) was then incubated with their respective blots at room temperature (with agitation) for an hour.

Excess primary antibody solution was then washed from the nitrocellulose membranes using 100 mL washing buffer (1X PBS and 0.1 % v/v Tween 20) with agitation and at room temperature for 10 minutes. This was then removed and repeated twice more. Western blots were then developed using a western blotting substrate (Pierce ECL Plus Western Blotting Substrate, Thermofisher Scientific) using 1: 1 ratio of Substrate A and B (3 mL total volume, per membrane); vortexing briefly prior to pipetting evenly across the membrane surface. Membranes were then incubated for 1 minute at room temperature prior to blotting and removal (via paper towel) of excessive substrate.

Data collection was performed using a Chemi G:Box F3 (Synegene) and GeneSys software (Version 3.2.2).

2.1.3 Estimation of protein molecular weights and oligomerisation states

2.1.3.1 Analytical size exclusion chromatography

The molecular weights of the recombinantly expressed CHFR proteins (FL-CHFR, Δ FHA1, Δ FHA2, Δ FHA3, FHA domain and C-terminal cysteine rich domain) were estimated using a size exclusion chromatography and molecular weight standards (Appendix 1, Section A1.5, Table A1.3).

A Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare) was equilibrated with 280 mL of gel filtration (GF) buffer (150 mM NaCl, 25 mM Tris-HCl (pH 7.0), 2 % (v/v) glycerol) using an ÄKTA Prime Plus chromatography system (GE Healthcare). A 5 mL sample injection loop (GE Healthcare) was filled with 10 mL GF buffer (via manually injected), followed by 0.5 mL of Gel Filtration Standard (resuspended in 0.5 mL MilliQ water, BioRad).

Using an ÄKTA Prime Plus chromatography system (GE Healthcare; 0.5 mL/ min), 15 mL of gel filtration buffer used to inject the sample from the loop to the column, proceeded by ~195 mL load of GF buffer (Appendix 1, Section A1.5, Figure A1.15).

All curves (UV absorbance and elution volumes) derived from SEC of proteins (FL-CHFR, ΔFHA1, ΔFHA2, ΔFHA3, FHA domain and C-terminal cysteine rich domain) and the molecular weight standard were then exported from the Unicorn System Control Software (Version 6.0) and analysed in Excel (Excel for Mac 2011, Version 14.7.2).

Per size exclusion dataset for each protein, the distribution of absorbance values (A_{280} , UV) was adjusted for the average and standard deviation per dataset and data was normalized using Excel's STANDARDIZE function (also known as the *standard score*) (Kreyszig, 1979):

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

Where X is the value to be normalized, μ the arithmetic mean and σ the standard deviation.

Initially, the average (or arithmetic mean) absorbance of the distribution was calculated using the AVERAGE function; where x is the final cell reference in the dataset:

$$\text{Average} = \text{AVERAGE}(\text{cell 1} : \text{cell } x) \quad (2)$$

In addition, the standard deviation of the distribution was calculated using the STDEV function; where x is the final cell reference in the dataset:

$$\text{Standard deviation} = \text{STDEV}(\text{cell 1} : \text{cell x}) \quad (3)$$

Finally, a normalized value from the distribution calculated (characterised by the calculated mean and standard deviation) was derived from Excel's STANDARDIZE function:

$$\text{Standardized value} = \text{STANDARDIZE} (\text{Cell2}, \text{STDEV}, \text{AVERAGE}) \quad (4)$$

Where Cell2 is in cell reference to the value to be normalized, STDEV is the cell reference to the standard deviation and AVERAGE is the cell reference to the calculated arithmetic mean.

To produce the calibration curve of molecular weight standards, the logarithm of proteins (to the base 10) was calculated using the Excel LOG function:

$$\text{Log}_{(10)} \text{ of molecular weight} = \text{LOG}(\text{Cell3}) \quad (4)$$

Where Cell3 is the cell reference to a value corresponding to the molecular weight (Da) for the molecular weight standard (Appendix 1, Section 1.5, Table A1.3). Log_{10} of molecular weight standards (Da, y axis) was then plotted against elution volumes per standard (x axis, mL). A linear trend line was then fitted, with the equation $y = -0.0438x + 8.164$ (Appendix 1, Section 1.5, Figure A.1.14).

Therefore, to calculate the molecular weights of the CHFR proteins, the elution volumes (mL, or x values) were substituted into the equation to obtain Y. The inverse log of the value of Y (molecular weight, log_{10}) was then calculated using the EXP Excel function:

$$\text{Molecular weight} = (10)^{\text{Cell4}} (5)$$

Where Cell4 is in reference to the cell containing the (equation derived) calculated \log_{10} molecular weight of selected CHFR protein. Finally, molecular weights (in Da) were converted to kDa.

The most probable molecular weight per CHFR protein was deduced by subtracting calculated molecular weights (as described above) from their predicted counterparts, per oligomeric state. The percentage difference in calculated molecular weights, with respective specific oligomeric state counterparts, was also calculated (Δm %). The oligomeric state with lowest difference in molecular weight in comparison to calculated molecular weight was deduced as the most probable. In addition, differences (kDa) between the predicted oligomer state and calculated counterpart were also deduced and tabulated.

2.1.3.2 Native PAGE analysis and western blots against CHFR proteins

To deduce the oligomerization state of CHFR proteins pooled across UV traces proceeding size exclusion chromatography and to assess homogeneity of purified protein samples, native PAGE analysis was conducted, proceeded by destaining (as previously described for SDS-PAGE gels, Section 2.1.2.3) and western blots against CHFR (Section 2.1.2.7).

The untagged CHFR C-terminus protein was diluted in SEC buffer (150 mM NaCl, 25 mM Tris-HCl (pH 7.0), 2 % (v/v) glycerol) across a range of concentrations (1.0 to 8.0 mg/ mL) to a final volume of 10 mL. In addition, a second set of native PAGE resolved gels consisted of CHFR (at 3 mg/ mL) and incrementally higher concentrations of BME added to samples (0 to 10 mM) to assess potential effects of reducing disulphide bonds upon C-terminal oligomerization.

The Strep(II)-tagged full-length CHFR protein was diluted in SEC buffer (150 mM NaCl, 25 mM Tris-HCl (pH 7.0), 2 % (v/v) glycerol) across a range of concentrations (0.5 to 2.0 mg/ mL) to a final volume of 10 mL.

Protein samples were resolved (within a cold room) using Native PAGE novel (4-16%) BisTris protein gels in 1X ice cold native page running buffer (25 mM Tris base, 192 mM glycine, adjusted to pH 8.3) with 5 uL loading dye (62.5 mM Tris-HCl (pH 6.8), 25% v/v glycerol, 1% (w/v) bromophenol blue) added per 10 uL protein sample. Prior to loading, samples were not heated. A molecular weight marker (1,236 to 20 kDa; NativeMark Unstained Protein Standard, Life Technologies) was also included.

Gels were resolved within an XCellSureLock Mini-Cell (Life Technologies) at (a constant) 40 V for 4 hours and destained over 24 hours (InstantBlue™, Expedeon); with residual destaining reagent removed with MilliQ water, prior to scanning.

In addition, selected proteins or assays resolved by native or SDS-PAGE were also subject to western blotting, using primary antibodies against either ubiquitin or CHFR, to significantly improve sensitivity in detection of proteins or polyubiquitin chain formation.

2.1.4 Prediction of coiled-coil motifs within the full-length CHFR protein

The use of a wide range of CHFR protein expression constructs (including FHA domain deletion constructs and individual domains) combined with size exclusion chromatography and prediction of molecular weights (and thus, oligomerization states) offers some insight into domain-mediated oligomerization within the full length CHFR protein. However, potential coiled-coils (CC) within domain linker regions (between both the N-terminal FHA domain and central RING domain, and the central RING domain and C-terminal cysteine rich domain) may be responsible for dimerization and potentially higher-order oligomerization remains unexplored.

Frequently responsible for protein oligomerization (McDonnell *et al.*, 2006), coiled-coils (CCs) are left-handed supercoils consisting of 2 to 7 intertwined, right-handed α -helices (Pauling *et al.*, 1951; Cohen and Parry, 1990) with enhanced structural stability exhibited via extensive helical packing (Cohen and Parry, 1986). Featured within selected filamentous (Crick, 1953) and globular (Pauling *et al.*, 1951) proteins; a heptad repeat (of seven residues; A, B, C, D, E, F, G) and interstitial apolar residue (at positions A and D) produce a stripe of apolar residues around the axis of individual α -helices (Pauling and Corey, 1953). Resulting ionic interactions and 'meshing' exhibited between individual α -helices within the CC interchains (Cohen and Parry, 1990) are attributed to an 18 (Crick, 1953) to 20° (Lupas *et al.*, 1991) axis incline per helix.

In order to identify potential coiled-coil motifs within domain linker regions of the full-length CHFR protein that may mediate protein dimerization or higher order oligomerization, two independent coiled-coil prediction tools were used.

Firstly, the COILS Server (Version 2.2.1; Lupas *et al.*, 1991) was used to predict coiled-coil regions outside of the N-terminal FHA, C-terminal cysteine rich and central RING domains of CHFR. Input protein sequences are initially compared to those within a local GenBank database, containing > 200 highly diverse proteins with (verified) two-stranded CCs (via a moving window) and scored on similarity (Lupas *et al.*, 1991). Calculated scores of target (input) sequences are then compared to their respective database standards and the probability of the protein adopting a CC fold (through the identification of heptad repeats within protein sequences) is calculated across a gliding window and provided as an output (Lupas *et al.*, 1991).

To maximise the coverage capacity of COILS in its identification of all possible heptad repeats within the protein sequences, all window widths (inclusive of 14, 21 and 28 residues) were selected; corresponding to a CC containing 4 to 5 heptad

repeats. For improved consistency, scoring capacity and resolution for CC and globular proteins (Lupas, Undated), the MITDK scoring matrix was selected.

In addition, probability of CC formation was calculated using the Paircoil2 (Version 1.0) server (Berger *et al.*, 1995; McDonnell *et al.*, 2006). Achieving fewer false-positive outputs in comparison to COILS, an incredibly diverse and extensive training database (including 1,371 protein chains) has also elevated the sensitivity of the algorithm employed. Through pairwise prediction, the target (input) and database sequences are tabulated; with the frequency of physically close residues (within a heptad repeat; or Pairwise Frequency Value, PFV) used to calculate the probability that specific residue pairs are specifically within a CC (Berger *et al.*, 1995). The *P*-score output is indicative of the percentage of CC-independent residues within the PDB subtracted from a better (respective) Paircoil score.

Within Paircoil2, minimum window search lengths of 21 and 28 were used in two independent predictions; with residues below a score of 0.025 predicted to be in a CC. Therefore, lower *P* scores imply a higher likelihood of CC formation (Berger *et al.*, 1995).

Using both COILS and Paircoil2, the FL-CHFR sequence was assessed for CC content. In addition, positive and negative controls with verified CC content (via X-ray structure) were also assessed, corresponding to the RNF8 (E3 ubiquitin ligase) (Campbell *et al.*, 2014; PDB: 4ORH; Reference Sequence: NP_003949.1) and lysozyme (Schoichet *et al.*, 1995; PDB: 2531L; GenBank: ACO37637.1) respectively.

2.1.5 Size exclusion chromatography in-line with small-angle X-ray scattering (SEC-SAXS) of CHFR proteins

In order to gain structural insight into interdomain interactions within the CHFR protein, previous attempts to crystalize the full-length protein (across a wide range

of screening conditions) for structure determination via X-ray crystallography proved unsuccessful; suggesting that a crystalline state of the full-length protein may be unattainable and therefore an unsuitable structure determination technique for the full-length protein.

Small angle scattering (SAS) of X-rays (SAXS) and neutrons (SANS) provides insight into the shape, low-resolution structure and structural parameters (such as radius of gyration, R_g , molecular mass, M_r , hydrated volume, V_p and maximum particle dimension, D_{max}) of biological macromolecules in solution (Svergun *et al.*, 2013). In contrast to crystallographic determination of biological macromolecular structure, SAS is advantageous in the omission of cryocooling and crystallographic sample preparatory steps.

Within a SAXS experiments, an incident (monochromatic) X-ray interacts with electrons distributed within the biological macromolecule, resulting in elastic X-ray scattering, recorded at a detector (Glatter, and Kratky, 1982) (Figure 2.1).

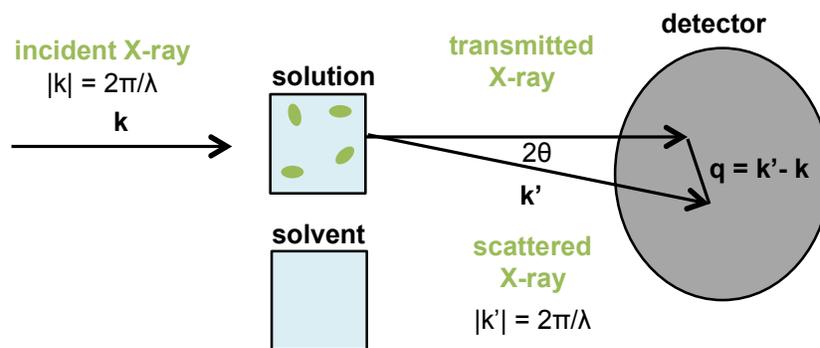


Figure 2.1 Schematic representation of a typical SAXS experimental set-up (adapted from Svergun *et al.*, 2013 and Skou *et al.*, 2014). Interactions occur between incident X-rays (k , with the modulus $|k| = 2\pi/\lambda$, where λ is the wavelength) and distributed electrons within the biological macromolecules (within solution) (Svergun *et al.*, 2013, Glatter, and Kratky, 1982). The elastically scattered X-rays ($k' = |k| = k$, therefore also $2\pi/\lambda$) are recorded at a 2D detector and dependent on the momentum transfer ($q = k' - k$) (Svergun *et al.*, 2013; Glatter, and Kratky, 1982). The scattering angle (2θ) between the incident X-ray beam and the scattered X-ray (therefore, direction of observation) (Svergun *et al.*, 2013, Glatter, and Kratky, 1982) is also indicated.

Net SAS intensities are obtained from subtracting measured solvent scattering patterns from that of the respective (measured) biological macromolecular solution scattering pattern, eliminating both instrument background and solvent scattering (Svergun *et al.*, 2013) (Figure 2.1). Proceeding normalization, background subtraction and instrument correction; the scattering pattern obtained ($I_{\text{total}}(q)$) is resultant of wave interference (Svergun *et al.*, 2013). The scattering intensities of incoming X-rays (with wavelength λ) scattered by biological macromolecules in solution, are dependent on the modulus of the momentum transfer ($|q|$) (Glatter, and Kratky, 1982; Svergun *et al.*, 2013); and therefore the variable q can be used to define the X-ray scattering of biological macromolecules observed in solution, where θ is half the angle of scattered X-rays:

$$q = 4\pi \sin \theta / \lambda \quad (6)$$

Following the initial purification of FL-CHFR, Δ FHA1, Δ FHA2 and Δ FHA3 proteins by SEC using a Superdex S200 column (Appendix 1, Section 1.4, Figures A1.3, A1.5 and A1.7, respectively), it was difficult to fully resolve peaks (corresponding to the eluted proteins) by SEC alone; whereby a small shoulder on the ascending chromatogram limb indicates some protein aggregates and/or impurities may be present. To resolve this, and to minimize the amount of aggregates prior to X-ray exposure, SAS using SEC in-line with SAXS (SEC-SAXS) was deemed most suitable technique.

In order to obtain both low-resolution models of the CHFR protein and deduce potential roles for individual (N-terminal FHA and C terminal cysteine rich) domains in mediating CHFR oligomerization in solution, SEC-SAXS was performed using FL-CHFR (0.43 mg/ mL), Δ FHA-2 (0.72 mg/ mL), C-terminal cysteine rich domain (8.69 mg/ mL) and FHA domain (4.83 mg/ mL) proteins. In addition, to obtain molecular weight information from the SAXS data, a BSA standard (10 mg/ mL) was also analysed by SEC-SAXS. Protein concentrations used for SEC-SAXS were approximate to: 100 divided by the protein molecular weight (kDa) (Graewert *et al.*, 2015); or:

$$\text{MW (kDa)} \times [\text{protein, mg/ mL}] \approx 100 \text{ (7)}$$

SEC-SAXS experiments using all 5 proteins (including a BSA control) were performed at beamline B21 by Dr Robert Rambo (principle beamline scientist) at Diamond Light Source (Oxford, UK). The in-line SEC-SAXS was performed using a 4.6 mL KW-402.5 analytical column (Shodex). Equipment was optimized for solution state SAXS experiments, whereby recorded data (on a Polatus 2M detector) was captured using a fixed camera length configuration (4.014 m) at 12.4 KeV; with a resolution range of 0.0038 to 0.42 Å⁻¹ (with a photon flux of ~10¹¹ per second) (Rambo, Undated) across 678 frames. 45 µL of each protein sample (BSA, FHA domain, C-terminal cysteine rich domain, full-length and ΔFHA-2 proteins; Appendix 1, Section 1.6, Figure A1.15, Panels A, B, C, D and E, respectively) was delivered via in-line Agilent HPLC (at 131.0 µL/ min) onto the SEC (pre-equilibrated in 150 mM NaCl, 25 mM Tris-HCl (pH 7.5), 2% (v/v) glycerol; adjusted to pH 7.0 with HCl).

2.1.5.1 Model independent analysis

2.1.5.1.1 Preliminary analysis and buffer subtraction

Analysis of unsubtracted SEC-SAXS data was performed using SCATTER (Version 1.0, Rambo, 2017); whereby selected frames before the protein elution peak was averaged and subtracted from selected frames within the protein elution peak (also averaged). Exported (buffer subtracted) output files were then analysed in PRIMUS (ATSAS Version 2.8.0; Franke *et al.*, 2017; Konarev *et al.*, 2006); whereby data was initially evaluated for correct buffer subtraction by plotting the log of scattering intensity (I(0), on an absolute scale) against momentum transfer (q); therefore assessing whether the measured scattering intensity remains positive and close to zero at higher q values.

2.1.5.1.2 Guinier Approximation

The Guinier approximation uses the Gaussian of normalized scattering curves from biological macromolecules, corresponding to the initial slope observed (Guinier, 1939; Svergun *et al.*, 2013):

$$I(q) \approx I(0)e^{-\frac{q^2 R_g^2}{3}} \quad (8)$$

Where R_g is the radius of gyration, which can be calculated plotting the momentum transfer squared (q^2) against the natural log of momentum transfer ($\ln I(q)$); identifying the zero intercept ($\ln I(0)$); within the Guinier Zone (or range) $0 < q < \frac{1}{R_g}$ (Guinier, 1939; Svergun *et al.*, 2013).

Guinier analysis was performed in PRIMUS (ATSAS Version 2.8.0; Franke *et al.*, 2017; Konarev *et al.*, 2006) by plotting the natural log of scattering intensity ($\ln I$) against the momentum transfer squared (q^2). By assessing the slope at low q values; aggregation, interparticle interference and multimerization can be also identified (Skou *et al.*, 2014; Svergun *et al.*, 2013).

2.1.5.1.3 Pair distance distribution

The pair-distance distribution ($P(r)$) function is an integral (obtained via fitting data within an inverse Fourier transformation, or IFT) defines the intensity of scattering observed between non-interacting particles; therefore describing the (average) distribution of particle orientations in solution (Porod, 1982; Svergun *et al.*, 2013):

$$I(q) = 4\pi \int_0^{D_{max}} P(r) \frac{\sin(qr)}{qr} dr \quad (9)$$

The function has an intergral limit corresponding to the particles maximum distribution (or maximal chord length, D_{max}) (Porod, 1982; Svergun *et al.*, 2013). PRIMUS (ATSAS Version 2.8.0; Franke *et al.*, 2017; Konarev *et al.*, 2006) was used to produce plots of r against $P(r)$.

2.1.5.1.4 Kratky plots

Attributing scattering behaviours to how well (or poorly) the proteins were folded in solution, quantitative analysis of Kratky plots ($q^2(I(q))$ plotted against q , Kratky, 1982) were performed in PRIMUS (ATSAS Version 2.8.0; Franke *et al.*, 2017; Konarev *et al.*, 2006).

2.1.5.1.5 Particle-mass estimation

Particle mass estimation was performed using two independent methods.

Firstly, the sample forward scattering intensity at zero angle ($I(0)$) previously determined via the Guinier approximation (Guinier, 1939) for both the protein standard (BSA) and CHFR FHA domain were substituted into the following equation, with their respective molecular weights (Graewert, 2015):

$$MW_{sample} = \frac{(I(0)_{sample} \times MW_{BSA})}{I(0)_{BSA}} \quad (10)$$

Secondly, the online SAXSMoW2 molecular mass estimation tool online (Fischer *et al.*, 2010) was also used. The software calculates molecular weights of proteins by deducing a linear relationship with apparent volumes V'):

$$V = A + BV' \quad (11)$$

Whereby V is the calculated, real volume (without hydration shell), A and B are linear and angular coefficients, respectively (Fischer *et al.*, 2010). The software then uses outputs of (A) V versus V' , (B) q_{max} versus A and (C) q_{max} versus B to calculate the estimated particle mass.

2.1.5.2 Model based analysis

Experimental X-ray structures of the CHFR N-terminal FHA domain are available via PDBs 1LGQ and 1LGP, corresponding to the dimeric protein without and the monomeric protein with tungstate, respectively (Stavridi *et al.*, 2002). However, differential construct lengths between X-ray experimental structures (spanning residues 14-124 or 14-128 (Stavridi *et al.*, 2002)) and the current in-solution, SEC-SAXS specific FHA domain protein (residues 13-180) make any theoretical scattering data from the X-ray structures (for example, derived using CRY SOL, Svergun *et al.*, 1995) incomparable to data from the in-solution (SEC-SAXS) counterpart protein.

Low resolution, *ab initio* shape models of the CHFR FHA domain protein (13-180) were therefore produced using DAMMIF, since the programme requires no *a priori* structural information to reconstruct the low-resolution shapes (Franke and Svergun, 2009). Specifically attributed to the programmes efficiency and guaranteed interconnectivity between final model beads (Franke and Svergun, 2009), DAMMIF was selected for model production (Franke and Svergun, 2009; Svergun, 1999).

DAMMIF (Franke and Svergun, 2009) (ATSAS Version 2.8.0; Franke *et al.*, 2017; Konarev *et al.*, 2006) was used to produce 20 *ab initio* molecular envelope models of the FHA domain dimer, with P2 (dimer) symmetry specified. DAMSUP (within ATSAS; Volkov and Svergun, 2003) was then used to align and average all 20 models (FHAdomai_1_*n*-1.pdb, where *n* is the model number) to a reference (initial) dummy model (model 1; FHAdomai_1_1-1.pdb). DAMAVER (within ATSAS; Volkov and Svergun, 2003) was then used to align the averaged models using co-ordinates specified in the output DAMSUP log file to generate a PDB model with most considerable shared bead position (or overlap) with other models.

All *ab initio* model figures were produced in Pymol, with individual model beads within particles represented by spheres or as a surface. Dimensions of the SAXS bead model were then measured and compared to the X-ray crystallography

experimental structure of both monomeric and dimeric FHA domains (PDBs 1LGQ and 1LGP; Stavridi *et al.*, 2002).

2.1.6 Ubiquitination assays

2.1.6.1 Expression and purification of His-tagged ubiquitin conjugating enzymes

2.1.6.1.1 Expression of His-tagged E2s

His6-Ubc13 His6-Mms2 and His6-UbcH5a were expressed and purified using an adapted version of established protocols (Ptak *et al.*, 2005; Lorick *et al.*, 2005).

Briefly, BL21 CodonPlus (DE3) RIL cells were transformed using Ubc13, Mms2 or UbcH5a expression vectors (Appendix 1, Section A1.3, Table 1.2); with 100 mL pre-cultures produced; as previously described for CHFR constructs (Section 2.1.2.1). Within a 1L conical flask, 1L of LB (supplemented with correct antibiotics), 10mL of pre-culture was added and incubated (37 °C, 5.844 x g; Infors HT, Multitron) to an optical density (at 600 nM, OD600) of 0.4 and induced with 0.4 mM IPTG. Cultures were incubated overnight and pellets retained.

2.1.6.1.2 Immobilized metal ion affinity chromatography of the (His-tagged) Ubc13 E2 ubiquitin-conjugating enzyme

A pellet was thawed on ice and resuspended in 20 mM Tris-HCl (pH 8.0), 10 mM MgCl₂, 1mM EDTA-NaOH (pH 8.0), 5 % (v/v glycerol) and 1 mM DTT; with a single (complete) protease inhibitor cocktail tablet added (EDTA-free, Sigma). The resuspended pellet was then sonicated on ice for 2 minutes (10 seconds on, 10 seconds off) with 40% power (Misonix Sonicator 3000, Cole-Parmer). Clarification of cell lysates was then performed by centrifugation using a JA 25.50 fixed angle

rotor at $43,588.584 \times g$ (19,000 RPM at r_{\max} of 108 mm) within an Avanti J-20 XP centrifuge (Beckman Coulter) for 45 minutes at 4 °C.

The retained supernatant was then filtered (0.45 μm Minisart, Sartorius) and adjusted to 8mM imidazole.

A 5 mL Histrap FF column (GE Healthcare) was washed (in 25mL MilliQ) and pre-equilibrated in 25 mL wash buffer (A); consisting of 50 mM NaPO_4 , 300 mM NaCl and 8 mM imidazole. The clarified lysate was then manually injected onto the column.

An ÄKTA Prime Plus chromatography system (GE Healthcare) was washed into the original filtered and de-gassed lysis buffer (Pump and Buffer A, 50 mM NaPO_4 , 300 mM NaCl and 8 mM imidazole) and a high-imidazole elution buffer (Pump and Buffer B, 50 mM NaPO_4 , 300 mM NaCl and 250 mM imidazole). The Histrap FF column (containing cell lysate) was then attached to the system and impurities washed off with Buffer A until the absorbance at 280 (A_{280}) was 0 mAu. Proteins were then eluted with a 80-100 mL gradient of 8 mM to 250 mM imidazole (at 3 mL/ min) via incrementally increasing the amount of buffer B added to buffer A (Appendix 1, Section A.17, Figure A1.25, Panel A). At approximately 1000 mAu, the UV peak corresponds to the His-6 tagged Ubc13 protein, with proteins resolved by SDS-PAGE analysis (Section 2.1.2.3; Appendix 1, Section A.17, Figure A1.25, Panel B).

His-tagged Ubc13 (17 kDa) was concentrated and selected fractions pooled and protein concentration determined as previously described (Section 2.1.2.2.3); substituting an extinction coefficient (corresponding to the His6-tagged Ubc13) of $22460 \text{ M}^{-1} \text{ cm}^{-1}$.

Concentrated proteins (with 5 % v/v glycerol added) were flash frozen in liquid nitrogen and stored at -80°C prior to usage.

2.1.6.1.3 Immobilized metal ion affinity chromatography of the (His-tagged) Mms2 E2 ubiquitin-conjugating enzyme

Immobilized metal ion affinity chromatography of the (His-tagged) Mms2 E2 ubiquitin-conjugating enzyme was performed as previously described for His-tagged Ubc13 (Section 2.1.6.1.2; Appendix 1, Section A1.7, Figure A1.26, Panel A); with proteins resolved by SDS-PAGE analysis (Appendix 1, Section A1.7, Figure A1.26, Panel B).

His-tagged Mms2 (16 kDa) was concentrated; with selected fractions pooled and protein concentration determined as previously described (Section 2.1.2.2.3); substituting an extinction coefficient (corresponding to the His6-tagged Mms2) of $23950 \text{ M}^{-1} \text{ cm}^{-1}$.

Concentrated proteins (with 5 % v/v glycerol added) were flash frozen in liquid nitrogen and stored at -80°C prior to usage.

2.1.6.1.4 Immobilized metal ion affinity chromatography of the (His-tagged) Ubch5a E2 ubiquitin-conjugating enzyme

Immobilized metal ion affinity chromatography of the (His-tagged) Ubch5a E2 ubiquitin-conjugating enzyme was performed as previously described for His-tagged Ubc13 (2.1.6.1.2; Appendix 1, Section A1.7, Figure A1.27, Panel A); with proteins resolved by SDS-PAGE analysis (Section 2.1.2.3; Appendix 1, Section A1.7, Figure A1.27, Panel B).

His-tagged Ubch5a (16 kDa) was concentrated; with selected fractions pooled and protein concentration determined as previously described (Section 2.1.2.2.3); substituting an extinction coefficient (corresponding to the His6-tagged Ubch5a) of $25440 \text{ M}^{-1} \text{ cm}^{-1}$.

Concentrated proteins (with 5 % v/v glycerol added) were flash frozen in liquid nitrogen and stored at -80°C prior to usage.

2.1.6.2 Ubiquitination assays and reaction component controls

Prior to characterization of CHFR- mediated polyubiquitination in the presence of both sets of E2 enzymes/ pseudo E2s (Ubc13: Mms2 and Ubch5a); a component control reaction was used to verify that individual reaction components are required for CHFR's E3 ubiquitin ligase activity.

The full (positive control) Ubc13 :Mms2 ubiquitin component control reaction (to a final 20 µL volume with MilliQ water, all final concentrations) contained 100 nM His-tagged E1 (Ube1, Ubiquigent), 5 mM MgCl₂, 5 mM DTT, 40 µM Tris-HCl (pH 7.5), 2.5 µM His-tagged Ubc13, 2.5 µM His-tagged Mms2, 1 µM Strep(II)-tagged full-length CHFR (1-664), and 100 µM untagged ubiquitin (Ubiquigent). An additional 4 assays were also performed as described above; omitting the E1 enzyme; both E2s (His-tagged Ubc13 and Mms2), FL-CHFR and ubiquitin and total volumes made to 20 µL. 2 µM ATP was added to start the reactions; incubated at 37°C for 1 hour.

Proceeding splitting the final 20 µL reaction volume into 2 x 10 µL tubes; SDS-PAGE analysis included both destaining gels (Section 2.1.2.3) and western blot analysis (Section 2.1.2.7); using a primary antibody against to ubiquitin to detect polyubiquitin chain formation, as previously described.

The full (positive control) Ubch5a ubiquitin component control reaction (to a final 20 µL volume with MilliQ water, all final concentrations) contained 100 nM His-tagged E1 (Ube1, Ubiquigent), 1 µM Strep(II)-tagged full-length CHFR (1-664), 5 mM MgCl₂, 5 mM DTT, 40 µM Tris-HCl (pH 7.5), 6.4 µM His-tagged Ubch5a and 100 µM untagged ubiquitin (Ubiquigent). An additional 4 assays were also performed as described above; omitting the E1 enzyme, His-tagged Ubch5a, FL-

CHFR and ubiquitin and total volumes made to 20 μ L. 2 μ M ATP was added to start the reactions; incubated at 37°C for 1 hour.

SDS-PAGE resolution of individual assay conditions (Section 2.1.2.3) and Western blot analysis (Section 2.1.2.7), using a primary antibody against to ubiquitin, was then performed to detect polyubiquitin chain formation, as previously described.

2.1.6.3 Verification of CHFR stability within ubiquitination assay conditions

To verify the stability of the CHFR protein within both the ubiquitination assay-specific (with final concentrations: 40 μ M Tris-HCl, pH 7.5, 1 μ M FL-CHFR, 5 mM MgCl₂, 5 mM DTT; to a 20 μ L final volume with MilliQ water) and E2Scan assay-specific (final concentrations: 50 mM HEPES-NaOH (pH 7.5), 5mM MgCL2, 5mM DTT, 1 μ M FL-CHFR; to a final volume of 20 μ L with MilliQ water) conditions; the CHFR protein was incubated at 37°C and SDS-PAGE analysis performed as previously described (Section 2.1.2.3).

Time points for ubiquitination assay-specific and E2Scan buffer conditions included: 0 (prior to incubation and the addition of ATP), 1, 2 and 3 hours.

2.1.6.4 FL-CHFR ubiquitination assays in the presence of either the Ubch5a or Ubc13: Mms2 E2 ubiquitin conjugating enzymes.

To initially characterize the rate of (Strep II) – full length CHFR- mediated polyubiquitin chain formation in the presence of the His-tagged Ubc13 E2 ubiquitin conjugating enzyme and His-tagged Mms2 pseudo E2 enzyme; a ubiquitination assay was performed within across time-course. This included 0 minutes (prior to incubation at 37 °C and the addition of ATP) and every 20 minutes thereafter, up to 3 hours. In addition, to ascertain whether non-specific K63 linked polyubiquitin chains form in the absence of the CHFR E3, an identical ubiquitination assay was performed in the absence of CHFR. To assess the rate of K63-linked chain

formation in the presence and absence of FL-CHFR, an additional K63 linked chain (0.5 μ L, 0.5 mg/ mL stock concentration, Ubiquigent) was also included.

This was then repeated to investigate UbcH5a-specific polyubiquitination chain formation in the presence and absence of FL-CHFR, with assay conditions identical to the initial reaction component control and those used for the Ubc13: Mms2 ubiquitination reactions.

SDS-PAGE resolution of individual assay conditions (Section 2.1.2.3) and Western blot analysis (Section 2.1.2.7), using a primary antibody against to ubiquitin, was then performed to detect polyubiquitin chain formation, as previously described.

2.1.6.5 Ubiquitination assays to assess whether the FL-CHFR FHA domain is essential for its activities as an E3 ubiquitin ligase

To assess whether the FHA domain is essential in CHFR's activities as an E3 ubiquitin ligase during either UbcH5a or Ubc13: Mms2 thioester conjugated chain formation; ubiquitination assays were performed as previously described (Section 2.1.6.4).

Assay conditions included: no E3 ubiquitin ligase (to control for non-CHFR specific chain formation), 1 μ M Strep(II)-tagged FL-CHFR (positive control, protein inclusive of FHA domain), 1 μ M Strep(II)-tagged CHFR Δ FHA2 and 1 μ M (untagged) CHFR C-terminus (394-664; as a negative control for CHFR RING-specific polyubiquitin chain formation); all with 40 μ M stock concentrations used.

Taking into consideration that within the Ubc13: Mms2-specific E2 assay, FL-CHFR polyubiquitin chain formation occurs at a higher rate within the first 20 minutes in comparison to reaction counterparts without the E3 ubiquitin ligase (Figure 2.16, Panel B), reaction times included 0 minutes (prior to incubation at 37 $^{\circ}$ C and the addition of ATP); and 5, 10 and 15 minutes post-incubation at 37 $^{\circ}$ C

(with 2 mM ATP). This ensured that reaction controls in the absence of FL-CHFR would specifically lack any K-63 linked polyubiquitin chains and therefore any chain formation was CHFR-specific. Therefore, samples were either resolved by SDS-PAGE analysis after 0 minutes (prior to addition of ATP and incubation at 37 °C); or after 5, 10 and 15 minutes at 37 °C (with 2 mM ATP).

SDS-PAGE resolution of individual assay conditions (Section 2.1.2.3) and Western blot analysis (Section 2.1.2.7), using a primary antibody against to ubiquitin, was then performed to detect polyubiquitin chain formation, as previously described.

2.1.7 Identification of new E2 enzymes responsible for CHFR-mediated polyubiquitin chain formation

2.1.7.1 E2Scan assay

Following the interaction between CHFR and the ubiquitin-loaded E2 enzyme, the E3 ubiquitin ligase catalyses the transfer of ubiquitin from the E2 to either a substrate conjugating it; or to itself via ubiquitination, forming an isopeptide bond with the ubiquitin protein. Whilst a selected number of E2 ubiquitin conjugating enzymes have been identified as ubiquitin conjugating enzymes responsible for CHFR-mediated substrate ubiquitination (Chaturvedi *et al.*, 2002; Kang *et al.*, 2002; Bothos *et al.*, 2003, Jung *et al.*, 2013), no other E2s responsible for coupling with CHFR have been identified.

In order to identify new E2 substrates for FL-CHFR-mediated polyubiquitin chain formation, an E2 Scan assay was used (according to manufacturer's instruction; (Anon, 2013)) to screen for new thioester conjugate enzymes responsible for isopeptide bond formation with the E3 ubiquitin ligase (E2Scan Kit Version 2, Ubiquigent).

The 96-well black ('v' bottom) plate contains 34 E2 ubiquitin conjugating enzymes (in duplicate; Appendix 1, Section A1.7, Table A1.4) at 0.05 nmoles/ well within a final 10 μ L volume of storage buffer (containing: 5mM DTT, 5% (v/v) glycerol, 5 mM $MgCl_2$ and 50 mM HEPES-NaOH (pH 7.5) (Appendix Table A2.3; (Anon, 2013))).

Prior to starting the assay, the E2Scan plate was thawed on ice for 20 minutes, then briefly centrifuged (1,000g for 1 minute at 4 $^{\circ}$ C, Allegra X-12R centrifuge, Beckman Coulter) to ensure that all contents had settled to the bottom of the wells; prior to carefully peeling open the plate.

Briefly, 5 μ L of ubiquitination reaction mix was added to one of the duplicate panels of E2 ubiquitin conjugating enzymes; containing 1.2 μ M His-tagged E1 (0.3 μ M final), 4 μ M FL-CHFR (WT, 1 μ M final), 400 μ M untagged ubiquitin (100 μ M final) and 1 X Ubiquitin Assay Buffer (UAB, prepared to the required final reaction mix volume from a 5 X stock using MilliQ water; 250 mM HEPES-NaOH (pH 7.5), 25 mM $MgCl_2$, 25 mM DTT).

To the second panel of (duplicate) E2 ubiquitin conjugating enzymes, an identical reaction mix was added, but with FL-CHFR omitted (and thus serving as a 'no-E3' control). This was used as a control to identify reaction conditions and E2 ubiquitin conjugating enzymes responsible for CHFR-specific polyubiquitin chain formation and reaction conditions with non-specific (non-CHFR dependent) polyubiquitin chain formation.

To start the reactions, 8 μ M ATP (20 mM) was added to all occupied wells and the plate briefly centrifuged at 1,000 g for 1 minute. The plate was then incubated at 37 $^{\circ}$ C for 3 hours; after which reactions were stopped by the addition of 5 μ L SDS-PAGE loading buffer. Samples were denatured and resolved by SDS-PAGE analysis, as previously described (Section 2.1.2.3) and western blots performed

using a primary antibody against ubiquitin to detect polyubiquitin chain formation, as previously described (Section 2.1.2.7).

2.1.7.2 CHIP (E3 ubiquitin ligase) positive control and buffer (E2 dependency) control

In addition to analysis of CHFR-mediated polyubiquitin chain formation, two positive control wells were used to verify correct E2Scan plate set-up and optimal reaction conditions for E3-mediated polyubiquitin chain formation. Containing 0.05 nmoles of the Ube2D4 E2 enzyme (in an identical E2 storage buffer to the other 34 E2 enzymes); 1 μ M CHIP E3 ubiquitin ligase (4 μ M stock) was added; with other reaction components identical to those previously described for the CHFR-positive panel (omitting FL-CHFR).

To check the E2 dependency of the ubiquitination reactions within the E2Scan experiment, two buffer controls (containing the complete reaction mix previously described, 1 μ M (final) FL-CHFR and no E2) were also included. It is expected that no polyubiquitin chain formation is to occur in the absence of an E2 ubiquitin conjugating enzyme.

To start the reactions, 8 μ M ATP (20 mM) was added to all occupied wells and the plate briefly centrifuged at 1,000 g for 1 minute. The plate was then incubated at 37 °C for 3 hours; after which reactions were stopped by the addition of 5 μ L SDS-PAGE loading buffer. Samples were denatured and resolved by SDS-PAGE analysis, as previously described (Section 2.1.2.3) and western blots performed using a primary antibody against ubiquitin to detect polyubiquitin chain formation, as previously described (Section 2.1.2.7).

2.1.7.3 CDC34/ CDC34B ubiquitination assay using untagged E2 enzymes

To verify whether CDC34 and CDC34B E2 polyubiquitin chain formation in the presence of the FL-CHFR E3 ubiquitin ligase was dependent on the E2 enzyme T7 tags and to attempt to reproduce the observed polyubiquitin chain formation within the E2Scan assay, an additional ubiquitination assay was set-up using identical reaction conditions (as detailed in Section 2.1.7.1); but substituting the E2 panel ubiquitin conjugating enzymes with either untagged CDC34 or untagged CDC34B (both Ubiquigent; at 0.05 nmoles/ well within a final 10 μ L volume of storage buffer (containing: 5mM DTT, 5% (v/v) glycerol, 5 mM MgCl₂ and 50 mM HEPES-NaOH (pH 7.5)); with FL-CHFR at 1 μ M (4 μ M stock).

As a positive control for both the reaction conditions and E2 enzyme storage buffer, UbcH5a (0.05 nmoles/ well within a final 10 μ L volume of storage buffer, as previously described) with FL-CHFR at 1 μ M (4 μ M stock) was also included.

To start the reactions, 8 μ M ATP (20 mM) was added and tubes briefly centrifuged at 1,000 g for 1 minute. Samples were then incubated at 37 oC for 1, 2 or 3 hours; after which reactions were stopped by the addition of 5 μ L SDS-PAGE loading buffer.

Samples were denatured and resolved by SDS-PAGE analysis, as previously described (Section 2.1.2.3) and western blots performed using a primary antibody against ubiquitin to detect polyubiquitin chain formation, as previously described (Section 2.1.2.7).

2.1.8 Phylogenetic analysis of E2 ubiquitin conjugating enzymes used in the E2Scan ubiquitination assay

2.1.8.1 Multiple sequence alignment of E2s

The E2Scan ubiquitination assay provides some insight into the specific E2 ubiquitin conjugating enzymes recognized by the CHFR proteins and responsible

for CHFR- mediated polyubiquitin chain formation *in vitro*. However, a potential relationship between biological similarity (or dissimilarity) concerning the sequences (and respective structures) of E2s and observed polyubiquitin chain formation remains unexplored.

Biological relationships between protein (and DNA) sequences can be explored using phylogenetics; involving the alignment of sequences derived from a set of homologues and the construction of a tree (Lesk, 2014). Modelling is performed by inferring either an explicit evolutionary process or via the classification of different datasets; whereby constructed trees can either depict homologues as originating from a single ancestral species (*rooted*) or infer relationships independent of any historical context (*unrooted*) (Lesk, 2014).

Through phylogenetic analysis of all 30 E2 enzymes via classification of E2 sequence datasets, biological relationships between different E2 conjugating enzymes can be cross-referenced with observable specificity of CHFR for E2s in polyubiquitin chain formation.

Sequences of all 30 E2 enzymes (corresponding to proteins used in the E2Scan assay) were imported into Jalview (Version 2.8.2; Waterhouse *et al.*, 2009) and aligned using Clustal Omega (Version 1.2.4; Sievers *et al.*, 2011); with 1 maximum guide tree iteration and a single iteration overall.

2.1.8.2 Constructing phylogenetic trees using Maximum Likelihood Estimation (MLE)

The Maximum Likelihood Estimation (MLE; Wilks, 1938) is a statistical method that can be used to construct phylogenetic trees. By calculating the probability of a mutation occurring, trees are constructed using a set of parameters to introduce the fewest number of mutations between different homologues (Lesk, 2014).

To perform phylogenetic analysis of E2 homologues using MLE, the exported multiple sequence alignment (MSA) of E2s (in a FASTA format, from JalView) was then converted to the correct format for PHYLIP (Version 3.695; Retief, 2000). An unrooted phylogenetic tree was constructed using the *sequence maximum likelihood* (proml) program in PHYLIP. The Jones-Taylor-Thornton (JTT) probability model (Jones *et al.*, 1992) was used to model change in amino acid sequence and a constant rate of variation assumed among sites; with a phylogenetic tree created using Drawtree (within the PHYLIP package).

2.1.8.3 Constructing phylogenetic trees using Bayesian inference

A second, independent method was also used to construct an unrooted phylogenetic tree to examine the biological relationship between the different E2 enzymes. MrBayes utilises the Bayesian method of statistical inference (or Bayesian probability) (Huelsenbeck and Ronquist, 2001); incorporated with a Markov chain Monte Carlo (MCMC) - based algorithm to sample posterior probabilities (Hastings, 1970) calculated for a generated range of phylogenetic trees (Tierny, 1994).

A phylogenetic tree for the 30 different E2 enzymes was generated using Bayesian inference and MCMC, the MSA of E2s (in a FASTA format, from JalView) via MrBayes (Version 3.2.3); selecting the 6 General Time Reversible (GTR) model for evolutionary data, default (protein) substitution modes specified and an assumed invariable rate of variation between different sites specified (Ronquist, 2011). The analysis was run for 10,000 generations; with diagnostics performed every 250 generations and the initial 250 trees discarded (Ronquist, 2011).

2.2 RESULTS

2.2.1 Dimerization of FL-CHFR is mediated by the C-terminal cysteine-rich domain and not the N-terminal FHA domain

To identify a potential role for the CHFR N-terminal FHA domain in mediating the dimerization of FL-CHFR and deduce whether other protein domains may have a role in CHFR oligomerization, 6 different CHFR proteins were expressed and purified (FL-CHFR, Δ FHA1, Δ FHA2, Δ FHA3, FHA domain and C-terminal cysteine rich domain; Figure 2.2, Panel A). In addition, analytical size exclusion chromatography was used in combination with protein molecular weight standards to produce a calibration curve and calculate the respective molecular weights of each CHFR protein (Figure 2.2, Panel B).

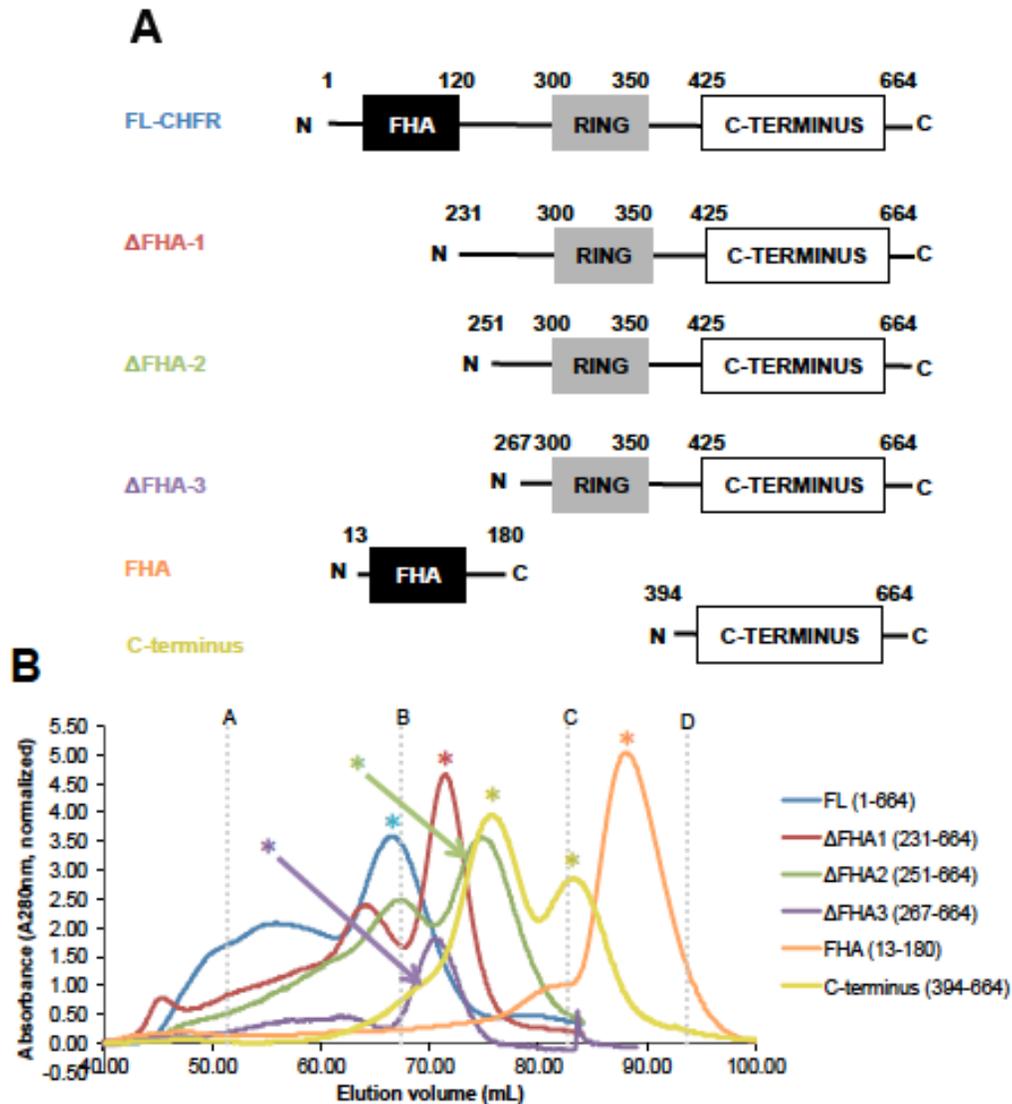


Figure 2.2 Calculation of the molecular weights and oligomerization of various CHFR constructs by size exclusion chromatography. (A) In order to deduce the oligomerization state of the full-length CHFR protein and whether protein dimerization is specifically attributed to segment-swapped dimerization previously reported in the crystal structure of the CHFR N-terminus alone (Stavridi *et al.*, 2002), a range of CHFR constructs; inclusive of the full-length protein, N-terminal deletion mutants (Δ FHA1, 2 and 3), FHA domain and c-terminal cysteine rich domain), were expressed and purified. (B) SEC of molecular weight standards (performed using identical experimental conditions used to purify CHFR proteins) enabled construction of a calibration curve (Appendix 1, Figures A1.15 and A1.16) and calculation of CHFR protein molecular weights and deduction of respective oligomerization states. Asterisks (*) indicate protein UV traces specifically associated with elution volumes of respective proteins. Elution volumes of specific molecular weight standards are indicated by the vertical grey lines, corresponding to A: Thyroglobulin (bovine, 670 kDa), B: γ -globulin (bovine, 158 kDa), C: Ovalbumin (chicken, 44 kDa) and D: Myoglobin (horse, 17 kDa).

Using the calibration curve produced from size exclusion chromatography of molecular weight protein standards (under identical conditions to CHFR proteins; Appendix 1, Table A1.3) and substitution of individual CHFR protein elution volumes (x) within the equation $y = -0.0438x + 8.164$; individual protein molecular weights were calculated (Table 2.2).

Table 2.2 Calculated molecular weights and deduced oligomerization states of CHFR proteins.

	Elution volume (mL)	Predicted MW (kDa) ^[1]		Calculated MW ^[2] (kDa); [Δm %] ^[4]	Deduced oligomeric state ^[5]
		<i>M</i>	<i>D</i>		
FL-CHFR	66.44	79.4	158.8	179.4 (+13.0 %)	D
ΔFHA1	71.39	61.1	122.2	108.9 (+10.9 %)	D
ΔFHA2	74.50	54.1	108.2	79.6 (-26.4 %)	D
ΔFHA3	73.27	52.3	104.6	90.0 (-14.0 %)	D
FHA	88.07	17.4	34.8	20.3 (+16.7 %)	M
C-terminus^[3]	(1) 75.68	30.4	60.8	(1) 70.7 (+16.3 %)	(1) D
	(2) 83.16			(2) 33.2 (+ 9.2 %)	(2) M

MW = molecular weight; *M* = monomer/ monomeric; *D* = dimer/ dimeric.

[1] Predicted molecular weights calculated using protein amino acid sequence within the ProtParam online tool (Gasteiger *et al.*, 2002).

[2] Molecular weights calculated using calibration curve produced using size exclusion chromatography and molecular weight standards; with known elution volumes of proteins (x) fitted to the curve $y = -0.0438x + 8.164$; where y is the calculated molecular weight.

[3] Two separate elution volumes, calculated molecular weights and deduced oligomerization states are included for the CHFR C-terminus, corresponding to each protein peak within the UV trace.

[4] [Δm %] indicates percentage differences with respect to the calculated (theoretical) mass (see [1]).

[5] Deduced oligomeric state corresponds to calculated MW with lowest percentage difference [Δm %] relative to predicted MW (see [1]).

The N-terminal FHA domain of CHFR appears to be monomeric in solution, with a single peak at 88.07 mL (5.03 mAu, normalized; Figure 2.2) corresponding to a (calculated) 20.3 kDa protein (Table 2.2) and a single band identifiable by SDS-PAGE analysis of selected SEC fractions (Appendix 1, Section 1.4, Figure A.11, Panel B).

In contrast, the calculated molecular weight (via molecular weight standards) of the full-length CHFR protein of 179.4 kDa (Table 2.2) is indicative that the protein is dimeric in solution, with a single peak observed and corresponding to an elution volume of 66.44 mL (Figure 2.2). Whilst dimeric FL-CHFR is observable via a low-concentration SEC elution (obtained from an initial 5 mL injection), a single band at 79 kDa across the elution peaks (Appendix 1. Section A1.4, Figure A1.3, Panel B) via SDS-PAGE analysis is indicative that quaternary structure disruption by heating (98°C for 5 minutes), denaturation (via SDS within the loading dye) and/or reducing agents results in lower order structures of the full-length CHFR protein being resolved. The expressed and purified FL-CHFR protein counterpart includes the N-terminal FHA and C-terminal cysteine rich domain; with a central RING domain also present (Figure 2.2, Panel A). Taking into consideration that the N-terminal FHA domain is monomeric under identical resolving conditions via SEC, it is most likely that the C-terminal cysteine rich domain is responsible for CHFR dimerization in solution.

Interestingly, the C-terminal cysteine rich domain of CHFR was eluted (by SEC) at two volumes (75.68 and 83.16 mL); corresponding to two distinct chromatogram peaks (Table 2.2; Figure 2.2, Panel B); with both monomeric and dimeric oligomerization states observed (33.2 and 70.7 kDa calculated molecular weights, respectively).

In addition, CHFR FHA domain deletion mutant proteins (Δ FHA1, -2 and -3) were eluted across individual UV peaks (Figure 2.2, Panel B) at volumes corresponding to molecular weights of their dimeric oligomerization states of 108.9, 79.6 and 90.0 kDa, respectively (Table 2.2); undetectable by SDS-PAGE analysis of selected SEC fractions (Appendix 1, Section 1.4, Figures A1.7, A1.5 and A1.9, respectively). This suggests the quaternary structure of Δ FHA proteins may be only observable at low concentrations (as identified by analytical SEC via a more initial dilute protein injection of 5 mL) and/or preserved under non-denaturing, non-reducing conditions.

Taking into consideration that the dimeric full-length CHFR protein and Δ FHA mutants contain the C-terminal cysteine rich domain (Figure 2.2 Panel A), it is most likely that dimerization of the Δ FHA and full-length CHFR proteins is dependent upon the C-terminal cysteine rich domain and not the N-terminal FHA domain.

2.2.2 The CHFR C-terminus forms a wide range of higher-order oligomers, whilst only dimeric full-length CHFR protein counterparts are observed

To assess whether concentration of the CHFR C-terminus had an effect on oligomerization state observed by SEC with molecular weight standards, multiple concentrations of the protein (1.0 to 8.0 mg/ mL) were resolved by native PAGE, ensuring any disulphide bond formation between protein monomers remained intact (Figure 2.3, Panel A). In addition, the effect upon oligomerization state of the CHFR C-terminus by the reduction of such bonds was assessed by the addition of BME (0 to 10 mM; Figure 2.3, Panel B). Gels were either destained or western blots performed using a primary antibody against the CHFR C-terminus to improve sensitivity.

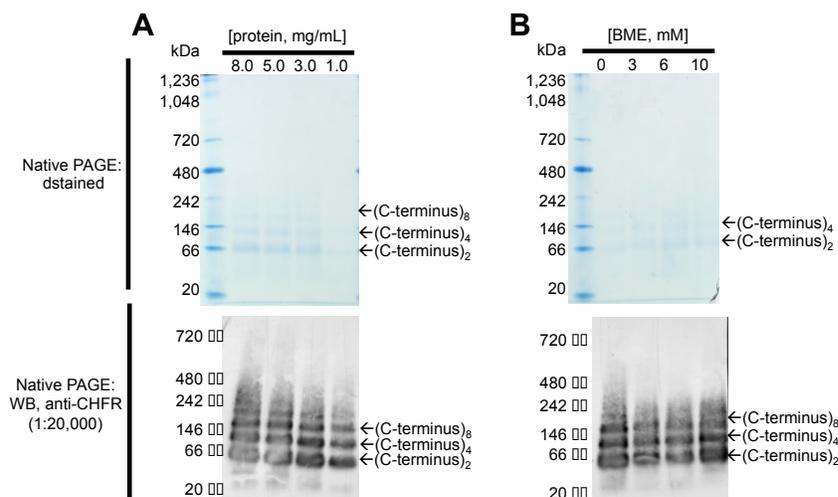


Figure 2.3 Native PAGE analysis of the CHFR C-terminal cysteine rich domain (394-664) reveals higher-order assembly complexes. (A) The purified CHFR C-terminus protein (concentrations ranging from 8.0 to 1.0 mg/ mL) was subjected to native PAGE under non-reducing and non-denaturing conditions; with gels either destained (top section) or blotted against a CHFR primary antibody (bottom section). (B) As described for A, but with increasing concentration of BME (0 to 10 mM).

In order to probe the oligomerization state of the CHFR C-terminus, native PAGE analysis under non-denaturing and non-reducing conditions offered the careful separation of intact CHFR C-terminal multisubunits, with a better resolution yielded and preservation of higher-order oligomers achieved, in comparison to SDS-PAGE and SEC analysis (Appendix 1, Section A1.4, Figure A1.14).

When the purified CHFR C-terminus was resolved on a native gel (4 – 16% gradient) and destained; the distribution of dimeric ((C-terminus)₂, ~60 kDa), tetrameric ((C-terminus)₄, ~120 kDa) and octameric ((C-terminus)₈, ~ 240 kDa) oligomers appears consistent across a wide range of protein concentrations (1.0 to 8.0 mg/ mL; Figure 2.3, Panel A). Interestingly, no lower-order and monomeric forms of the CHFR C-terminus (~ 30 kDa) are observed; either by destaining the native gels (Figure 2.3, upper panels) or via (the more sensitive) western blot analysis of native gels using a primary antibody against the CHFR C-terminus (Figure 2.3, lower panels). In addition, identical native PAGE conditions used to resolve the CHFR C-terminus in the presence of an increasing amount of (BME 0 to 10 mM, all at 3mg/ mL; Figure 2.3 Panel B), to reduce any disulphide bond formation at the CHFR C-terminus, has no effect on oligomer distribution previously observed (Figure 2.3, Panel B compared against Lane 3 of Panel A). Once again, no monomeric C-terminal proteins were observed either by destaining native gels (Figure 2.3 upper panels) or by western blot (Figure 2.3, lower panels).

In contrast, monomeric forms of the CHFR C-terminus (a single band at ~30kDa) were observed proceeding SDS-PAGE analysis of the purified protein under denaturing (via the SDS within the running buffer and heating samples prior to loading) and reducing conditions (Appendix 1, Section 1.4, Panel B). However, 2 distinct forms of the CHFR C-terminus as both monomer and dimer may be present within the SEC chromatogram (Appendix 2.1, Figure A2.14, Panel A; Figure 2.2, Panel B); whereby distinct UV peaks at 75.68 and 83.16 mL correspond to dimeric (70.7 kDa) and monomeric (33.2 kDa) C-terminal proteins, respectively (Appendix 1, Section 1.4, Figure A1.14, Panel B). In addition, a slightly higher

distribution of the dimeric C-terminus is present, in comparison to the monomeric counterpart (eluted with normalized absorbencies of 3.94 and 2.85 mAu, respectively; Appendix 1, Section 1.4, Figure A1.14, Panel B).

Whilst concentration dependent oligomerization of the CHFR C-terminus cannot be inferred within the explored range via native PAGE analysis (8.0 – 1.0 mg/ mL; Figure 2.3, Panel A), SEC in conjunction with MW standards (Figure 2.2, Panel B) clearly indicates CHFR C-terminal monomers can be resolved from dimeric counterparts at lower concentrations (via a 5 mL injection of protein) by SEC.

Taken together, these results indicate that disulfide bond formation does not have a role in the formation of CHFR C-terminal higher-order oligomerization states; whereby identical distributions of di-, tetra- and octomeric C-terminal protein are observed in the presence of up to 10 mM BME (all at 3mg/ mL; Figure 2.3, Panel B). However, at lower concentrations, monomeric and dimeric forms of the CHFR C-terminus (Figure 2.2, Panel B) are identifiable; suggesting some concentration dependency in CHFR C-terminal oligomerization states at < 1 mg/ mL. In addition, formation of di-, tetra- and octomeric CHFR C-terminus oligomers is temperature sensitive; with complete destruction of quaternary structures observed by heating at 98°C for 5 minutes (as observed in SDS-PAGE resolved samples).

To ascertain the potential effect of CHFR C-terminal oligomerization upon higher-order structure formation in the full-length CHFR protein, native PAGE analysis was also performed using strep(II)-tagged, FL-CHFR (without BME) at a range of concentration (0.5 to 2.0 mg/mL; Figure 2.4).

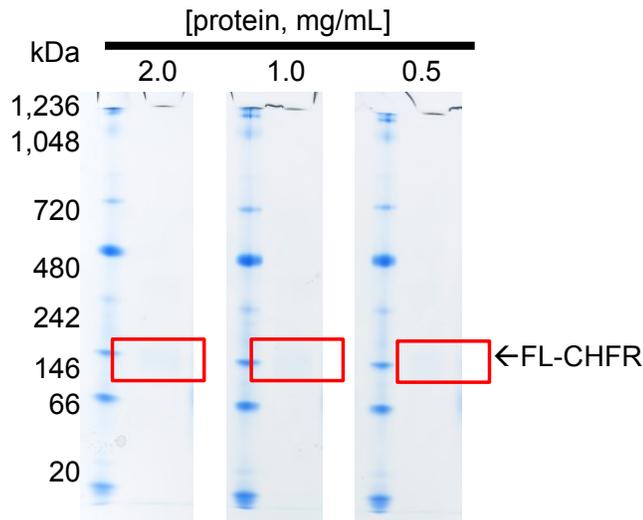


Figure 2.4 Native PAGE analysis of Strep(II) full-length CHFR protein (1-664) reveals a single dimeric species. (A) The purified FL-CHFR protein (concentrations ranging from 0.5 to 2.0 mg/ mL) was subjected to native PAGE under non-reducing and non-denaturing conditions; with gels destained.

Native SDS-PAGE analysis of the full-length CHFR protein under non-denaturing and non-reducing conditions offered the careful separation of the intact full-length CHFR dimer, with a better resolution yielded and preservation of higher-order oligomers achieved, in comparison to SDS-PAGE analysis (Appendix 1, Section A1.4, Figure A1.3).

When the purified FL-CHFR protein was resolved on a native gel (4 – 16% gradient) and destained, a single band corresponding to the dimeric protein (~158 kDa, Table 2.2) could be observed. No lower-order and monomeric forms of the full-length protein (79 kDa) are observed (Figure 2.4), even at a very low protein concentration (0.5 mg/ mL).

Moreover, a single peak resolved by SEC and corresponding to the FL-CHFR protein (Figure 2.2) is indicative that the full-length protein remains dimeric even at very low concentrations; whereby a 5 mL injection of protein was typically pooled across up to 5 individual 3 mL SEC fractions. In addition to the FL-CHFR protein, Δ FHA CHFR proteins (Δ FHA1, -2 and -3) were deduced as dimeric by SEC with molecular weight standards (Figure 2.2, Panel B; Table 2.2); inferring that the CHFR-C terminus is most likely to be responsible for protein dimerization.

The wide range of C-terminal dimeric, tetrameric and octomeric oligomerization states observed by native PAGE (Figure 2.3) are however absent in the resolved full-length counterpart, even at high (and comparable) 1.0 and 2.0 mg/ mL protein concentrations (Figure 2.4).

2.2.3 CHFR dimerization is not the result of coiled-coil motifs outside the FHA, central RING or C-terminal cysteine rich domains

The use of SEC coupled with molecular weight standards and native PAGE analysis using a wide range of CHFR constructs offers some insight into domain-mediated oligomerization of CHFR. However, potential secondary structural content within domain linker regions (between both the N-terminal FHA domain and central RING domain, and the central RING domain and C-terminal cysteine rich domain; Figure 2.2, Panel A) responsible for dimerization and potentially higher-order oligomerization remains unexplored.

In order to identify coiled-coil motifs within linker regions that may be responsible for oligomerization of the full-length CHFR protein, two independent coiled-coil prediction techniques (including the COILS server (Version 2.2; Lupas *et al.*, 1991; Figures 2.5) and Paircoil2 (McDonnell *et al.*, 2006; Berger *et al.*, 1995; Figure 2.6) were used to predict coiled-coil content.

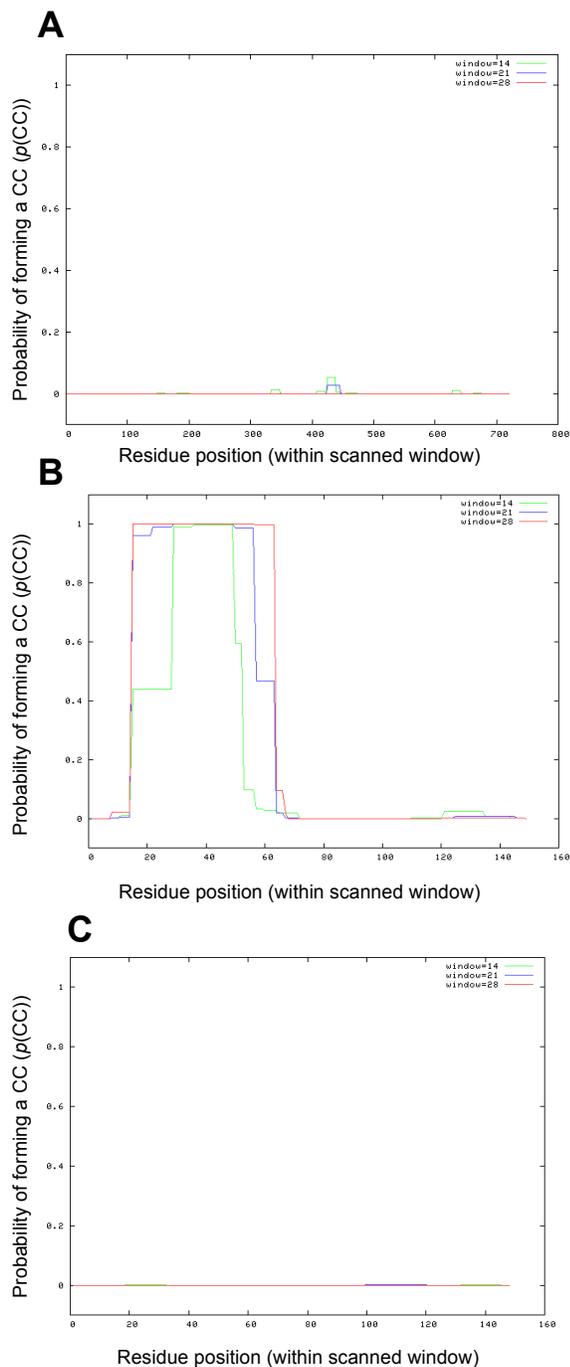


Figure 2.5 Prediction of FL-CHFR coiled-coil (CC) structural content using the COILS server (Version 2.2.1; Lupas *et al.*, 1991) indicates that CHFR is very unlikely to contain any coiled-coils. (A) Calculated using the FL-CHFR sequence, exceptionally low probability scores of CC formation ($p(\text{CC}) \leq 0.020$) within all three windows (14, 21 and 28, respectively) were calculated. (B) Higher probabilities of CC formation were calculated using the RNF8 positive (E3) control sequence ($0.222 \leq p(\text{CC}) \leq 1.000$); corresponding with the majority of respective CC content confirmed via the experimental obtained using X-ray crystallography (Campbell *et al.*, 2012; PDB: 4ORH). (C) Probabilities of CC formation calculated using the lysozyme negative control are indicative of no coiled-coil content ($p(\text{CC}) \leq 0.020$); with results corroborating with the lysozyme X-ray structure (Schoichet *et al.*, 1995; PDB: 2531L). 14, 21 and 28 residue windows are indicated by the green, blue and red traces, respectively.

Using the sequence-wide scanning profiles employed within the COILS server (Lupas *et al.*, 1991), exceptionally low probability scores were calculated ($p(\text{CC}) \leq 0.020$) for predicted CC structural content within the lysozyme negative control sequence (Figure 2.5, Panel C). Since this is strongly comparable to the FL-CHFR predicted CC content (Figure 2.5, Panel A; also $p(\text{CC}) \leq 0.020$), this is indicative that CC formation within the FL-CHFR protein is very unlikely.

As expected, the positive RNF8 control sequence for prediction of CC structural content using COILS (Figure 2.5, Panel B) resulted in the calculation of high probabilities of CC formation ($0.249 \leq p(\text{CC}) \leq 1.000$) within regions comparable to that of the reported experimental structure (Campbell *et al.*, 2012).

Using the pair-wise residue correlation method employed by the Paircoils2 server (McDonnell *et al.*, 2006, Berger *et al.*, 1995), exceptionally high P -scores were calculated ($p\text{-score} \geq 0.51904$) for predicted CC structural content within the lysozyme negative control sequence (Figure 2.6, Panel C). Since this is strongly comparable to the FL-CHFR predicted CC content (Figure 2.6, Panel A; (≥ 0.37839)), CC formation within the FL-CHFR protein is very unlikely.

As expected, the positive RNF8 control sequence for prediction of CC structural content using Paircoil2 (Figure 2.6, Panel B) resulted in the calculation of high probabilities of CC formation (corresponding to low P -scores, $0.00538 \leq p\text{-score} \leq 0.01974$) within regions comparable to that of the reported experimental structure (Campbell *et al.*, 2012).

For qualitative comparison of the accuracy regarding CC predication within the RNF8 protein using either the COILS or Paircoil2 servers, all residues within the high probability ($p(\text{CC}) \leq 0.020$) or high P -score (<0.0250) range were evaluated against the protein experimental structure (Figure 2.7).

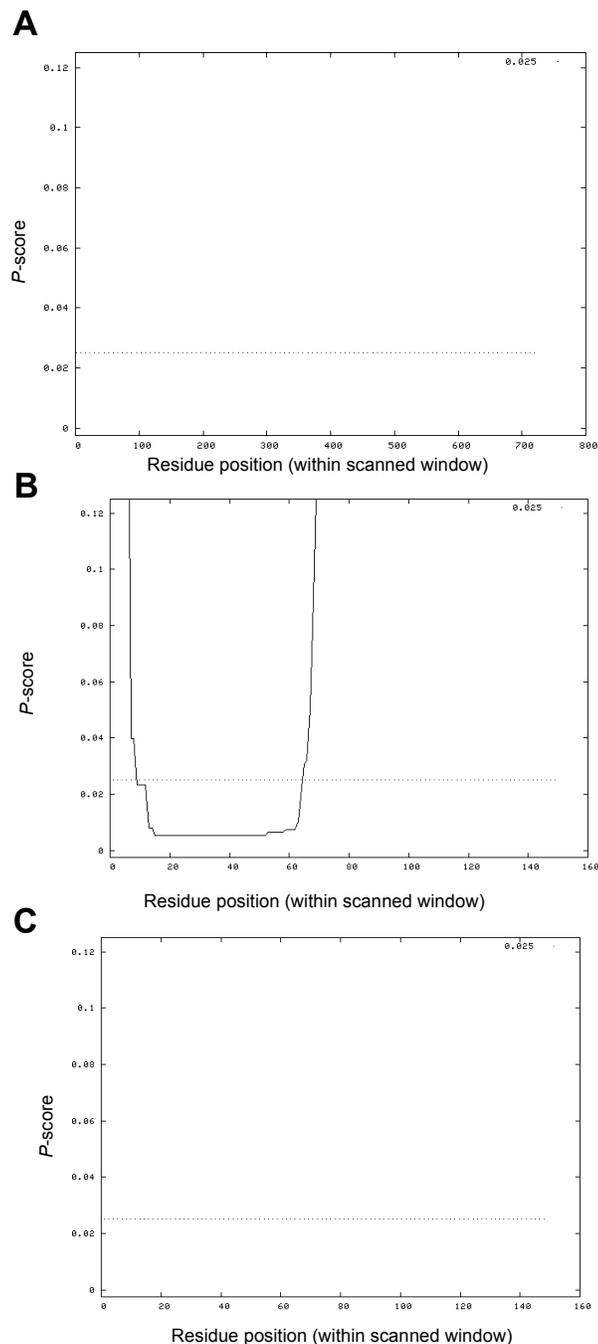


Figure 2.6 Prediction of FL-CHFR, coiled-coil (CC) structural content using the Paircoil2 server (Version 1.0; McDonnell *et al.*, 2006; Berger *et al.*, 2005) (with 21 and 28 residue minimum search window) indicates that CHFR is very unlikely to contain any coiled-coils. (A) Calculated using the FL-CHFR sequence, exceptionally high P -scores across all residue positions (≥ 0.37839) exceed the specified cut-off (< 0.025), indicative that the protein is unlikely to contain any CCs. (B) Significantly lower P -scores were calculated using the RNF8 positive (E3) control sequence ($0.00538 \leq p\text{-score} \leq 0.01974$); corresponding with majority of respective CC content confirmed via the experimental obtained using X-ray crystallography (Campbell *et al.*, 2012; PDB: 4ORH). (C) Probabilities of CC formation calculated using the lysozyme negative control are indicative of no coiled-coil content ($p\text{-score} \geq 0.51904$); with results corroborating with the lysozyme X-ray structure (Schoichet *et al.*, 1995; PDB: 2531L)

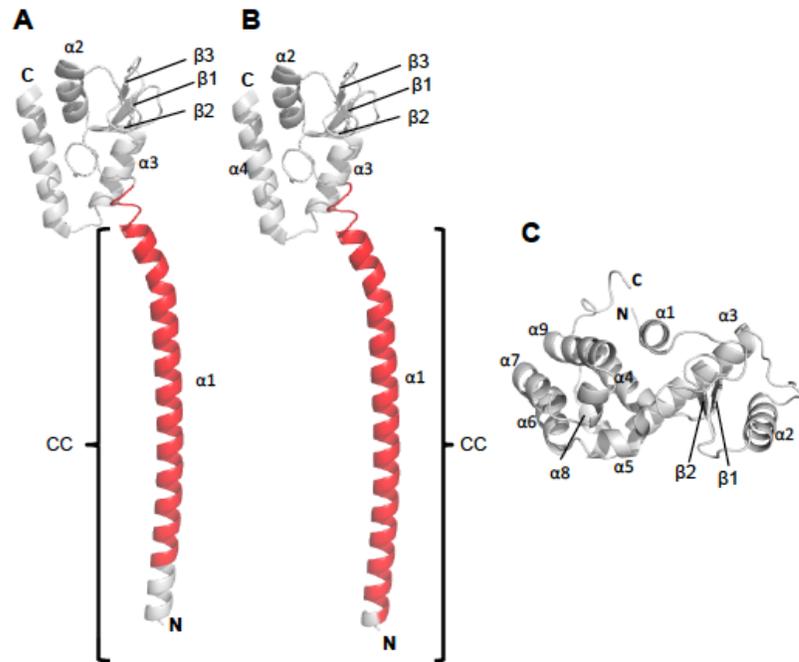


Figure 2.7 Comparison of (A) COILS (Lupus *et al.*, 1991) and (B) Paircoil2 (McDonnell *et al.*, 2006; Berger *et al.*, 1995) coiled-coil predictions made using the RNF8 protein (Campbell *et al.*, 2012) and (C) lysozyme (Shoichet *et al.*, 1995) experimental structures. Residues within the high probability ($p(CC) \leq 0.020$) or high P -score (<0.0250) range of the (A) COILS or (C) Paircoil2 predictions, respectively, were mapped onto the experimental structure of the RNF8 protein; as shown in red. CC: full length of coiled-coil. (C) using both prediction tools, no CC regions were predicted within the lysozyme protein.

Whilst both methods accurately detected the majority of RNF8 CC structural features, both predicted a false-positive region towards the C-terminus of the first alpha helix ($\alpha 1$, Figure 2.7, Panels A and B), spanning residues M392 to N399. In addition, both a false-negative region is evident at the N-terminus of $\alpha 1$; improved detection via Paircoil2 (omitting experimentally validated CC residues P342 to F334 from CC prediction), in comparison to COILS (omitting experimentally validated residues P342 to L350 from CC prediction; Figure 2.7). Since this corresponds to the equivalent of at least one heptad repeat, the Paircoils2 server produces fewer false positives, as demonstrated via the technique evaluation here (versus the COILS server) and by McDonnell *et al* (2006).

Taken together, CC prediction using both Paircoil2 and COILS servers indicates that it is very unlikely that the FL-CHFR protein contains any CC; with expected

outcomes via lysozyme and RNF8 CC predictions serving as valid (and strongly supporting) negative and positive controls (respectively) when compared to their experimental structures.

2.2.4 Model-independent analysis of the CHFR FHA domain indicates an extended and dimeric protein in solution

In order to obtain low-resolution structural information and parameters regarding FL-CHFR, Δ FHA-2, C-terminus and FHA domain proteins, SAXS were performed in-line with SEC; ensuring data was from monodisperse, non-aggregated samples.

The original SEC chromatograms (Appendix 1, Section A1.6, Figure A.15) were assessed for quality of resolution between individual protein peaks; whilst unsubtracted SEC-SAXS data (analysed in SCATTER; Rambo, 2017) was used to assess signal intensity of eluted proteins (indicated by peaks), relative to buffer (pre-protein elution, or signal: noise). Poor resolution between two peaks (and therefore different protein oligomeric states combined within a single elution) within the original chromatograms from SEC performed using the C-terminus and FL-CHFR proteins (Appendix 1, Figure A1.15, Panels C and D respectively) can be observed. In addition, three individual protein species identified via 3 different peaks preceding Δ FHA-2 SEC (Appendix 1, Section A1.6, Figure A1.15, Panel B) indicate protein oligomerization and/or degradation. In addition, the incredibly low integrated intensities observed across all 678 frames of data collected for both FL-CHFR and Δ FHA-2 proteins (potentially due to low initial concentrations injected; 0.43 and 0.72 mg/ mL respectively) and high noise also deemed the data sets was also observed (Appendix 1, Section A1.6, Figure A1.16). Taken together, data acquired from SEC-SAXS experiments using FL-CHFR, Δ FHA-2 and C-terminal proteins were omitted from additional analysis. In contrast, data obtained from SEC-SAXS analysis of both the BSA control and CHFR FHA domain was deemed suitable for additional data analysis (Table 2.3; Appendix 1, Section 1.6, Figure A1.17).

Table 2.3 Data collection and scattering parameters for SEC-SAXS analysis of BSA and CHFR FHA domain proteins.

Sample	BSA[~]	FHA
Data collection parameters		
Instrument*	B21	
Beam geometry*	1 x 8 mm	
Wavelength (Å)*	1.00	
q range (Å ⁻¹)*	0.0022 - 0.4200	
Exposure time per frame (sec)*	3.0	
Concentration (mg/ mL)	10.0	4.83
Temperature (K)*	266	
SEC specifications		
Flow rate (μL/ min)*	131.0	
Volume loaded (μL)*	45.0	
Structural parameters		
$I(0)$ (AU) [from $P(r)$]	0.14 ± 9.1e-5	0.012 ± 3.2e-5
R_g (Å) [from $P(r)$]	28.58 ± 0.12	24.54 ± 0.33
$I(0)$ (AU) [from Guinier]	0.14 ± 7.5e-5	0.012 ± 3.3e-5
R_g (Å) (from Guinier)	28.52 ± 0.02	24.51 ± 0.11
D_{max} (Å) [from $P(r)$]	90.00	85.00
Porod volume estimate (10 ³ Å ³)	103.00 ± 20	41.00 ± 20
Dry volume calculated from sequence (10 ³ Å ³) [§]	83.876	21.052
Molecular mass determination		
Molecular mass M_r from SAXSMoW [^] (KDa) [Δm %]	68.307 (+2.171%)	25.923 (-25.500%)
Calculated monomeric M_r from sequence (KDa)	66.500	17.398
Calculated dimeric M_r from sequence (KDa)	NA	34.796
Software employed		
Primary data reduction	Performed at the beamline	
Data processing*	PRIMUS (ATSAS)	
<i>Ab initio</i> analysis	DAMMIN (including DAMSUP and DAMAVER; ATSAS)	
Three-dimensional graphics representation*	Pymol*	

* Identical parameters for all samples; AU: arbitrary units.

[~]: Control protein sample; [§] calculated using Property Calculator (Chazan, Undated).

[^] Calculated using the online SAXSMoW2 tool (Hannes *et al.*, 2010).

NA: not applicable.

[Δm %] indicates percentage differences with respect to the calculated (theoretical) mass.

Unsubtracted SEC-SAXS data from the CHFR FHA domain protein sample was analysed using SCATTER; whereby a monodisperse and non-aggregated protein sample was identified via identification of single peak corresponding to the eluted protein (Appendix 1, Section 1.6, Figure A1.17). Frames corresponding to buffer (or *background*) preceding the protein elution peak (frames 1 to 250) were then averaged and subtracted from the averaged frames corresponding to the main eluted FHA domain protein peak (frames 380 to 430).

The output file, consisting of the buffer subtracted scattering intensity curve (Appendix 1, Section 1.6, Figure A1.18), was then analysed in PRIMUS (within the ATSAS software suite, Version 2.8.0; Franke *et al.*, 2017). Correct buffer subtraction within the experimental scattering pattern (Figure 2.8) was verified by a positive scattering intensity (close to zero) with higher q values (Skou *et al.*, 2014) (Appendix 1, Section A1.6, Figure A1.18).

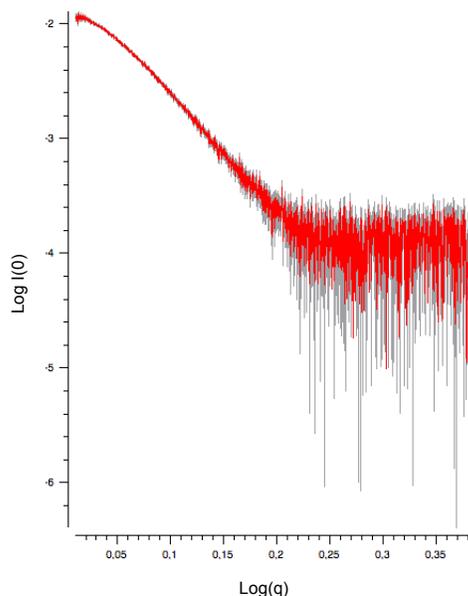


Figure 2.8 Experimental scattering pattern from the CHFR FHA domain, with buffer subtraction.

Production of a Guinier curve ($\ln(I)$ against q^2) and analysis of the low q region is indicative of no downturn of values and good linearity; therefore confirmation of a

non-aggregated, monodisperse protein sample, without interparticle interference (Appendix 1, Section A1.6, Figure A1.19).

A similar process of buffer subtraction was also performed for the BSA protein control, whereby frames corresponding to the buffer intensities (frames 1 to 280) were averaged and subtracted from frames corresponding to the protein peak (frames 390 to 400) (Appendix 1, Section A1.6, Figure A1.21).

The output file, consisting of the buffer subtracted scattering intensity curve (Figure 2.9), was then analysed in PRIMUS (within the ATSAS software suite, Version 2.8.0; Franke *et al.*, 2017). Initially plotted as the log of scattering intensity ($I(q)$) against momentum transfer (q , at an absolute scale); correct buffer subtraction was verified by a positive scattering intensity (close to zero) with higher q values (Appendix 1, Section 1.6, Figure A1.20).

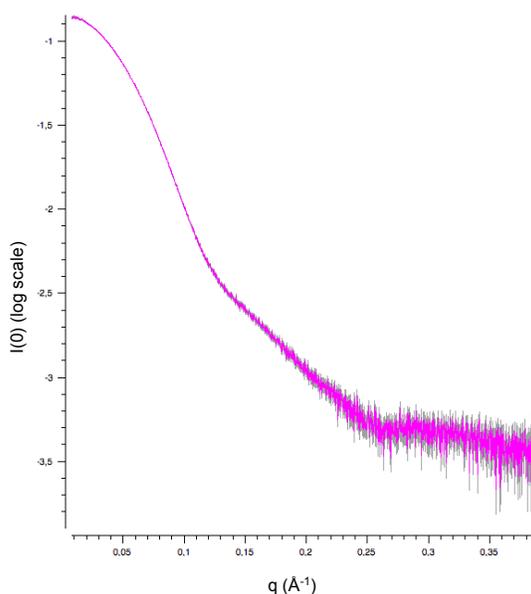


Figure 2.9 Experimental scattering pattern from the BSA protein control, with buffer subtraction.

Production of a Guinier curve ($\ln(I)$ plotted against q^2) and analysis of the low q region is indicative of no downturn of values and good linearity; with sample confirmed as monodisperse and absent of interparticle interactions and aggregation (Appendix 1, Section A1.6, Figure A1.22).

The pair distance distribution function ($P(r)$) was then used to assess whether the CHFR FHA domain or BSA control proteins were well-folded in solution, aggregated or exhibiting interparticle repulsion using the PRIMUS (Figure 2.10).

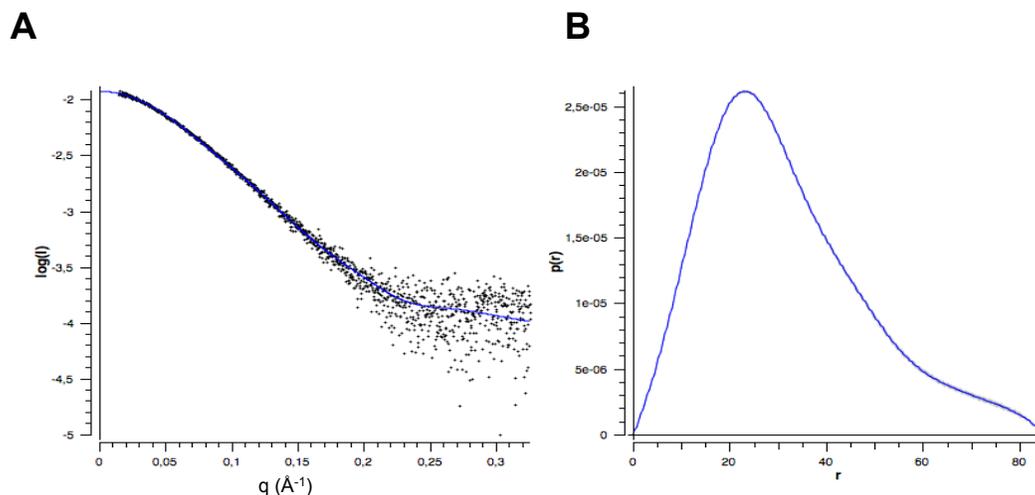


Figure 2.10 SAXS of the CHFR FHA domain. (A) The scattering intensity profile (log of intensity against momentum transfer) is shown for the CHFR FHA domain, with line fitted across high and low q regions (blue). (B) The frequency distribution of particle distances depicted within the $P(r)$ plot was used to calculate the R_g and D_{\max} of 24.54 ± 0.33 and 85.00 \AA , respectively.

The elongated tail region of the peak and non-gradual approach of $P(r)$ values towards zero (at higher r region; Figure 2.10, Panel B) may be indicative of an extended conformation. Fitting the SAXS data using the inverse Fourier Transform (IFT) within the same range as the initial Guinier analysis produced R_g and D_{\max} values of 24.54 ± 0.33 and 85.00 \AA , respectively. Good agreement (within error, Table 2.3) is exhibited between R_g and $I(0)$ values derived from both the Guinier analysis (Appendix 1, Section 1.6, Figure A1.19) and the pairwise distribution function (Figure 10).

In contrast, the atom-pair distance distribution exhibited within the $P(r)$ plot for the BSA protein control indicates a well-folded and globular protein in solution; whereby a gradual smoothing of the curve as $P(r)$ values approach zero (at high r values) can be observed (Figure 2.11, Panel B).

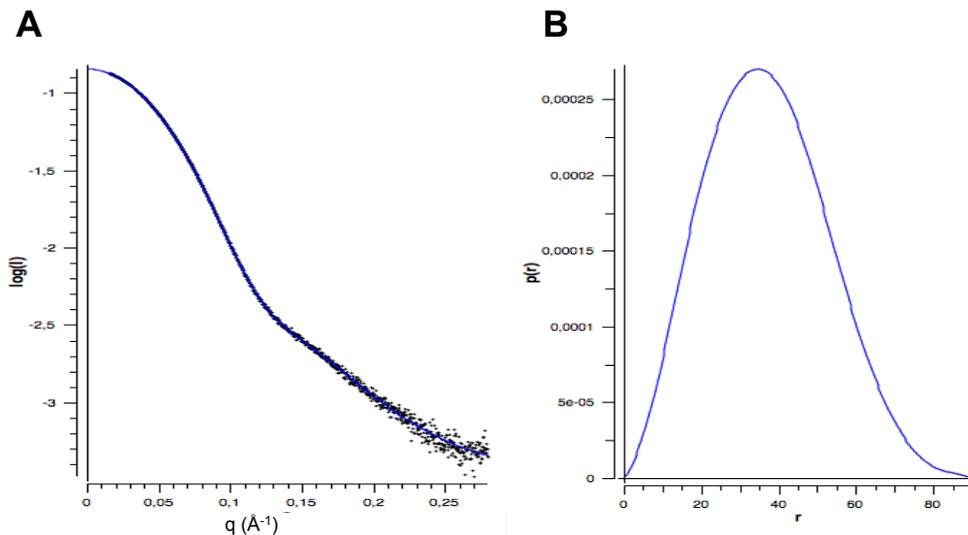


Figure 2.11 SAXS of the BSA protein control. (A) The scattering intensity profile (log of intensity against momentum transfer) is shown for BSA, with line fitted across high and low q regions (blue). (B) The frequency distribution of particle distances depicted within the $P(r)$ plot was used to calculate the R_g and D_{\max} of 28.58 ± 0.12 and 90.00 \AA , respectively.

Good agreement (within error, Table 2.3) is exhibited between R_g and $I(0)$ values derived from both the Guinier analysis (Appendix 1, Section A1.6, Figure A1.22) and the pairwise distribution function (Figure 2.11).

Quantitative analysis of a Kratky plots (plotted as $q^2 I(q)$ against q ; Kratky, 1982) for both the BSA control and the CHFR FHA domain proteins (Figure 2.12, Panels A and B, respectively) exhibit well-defined bell shaped curves within the mid- q range.

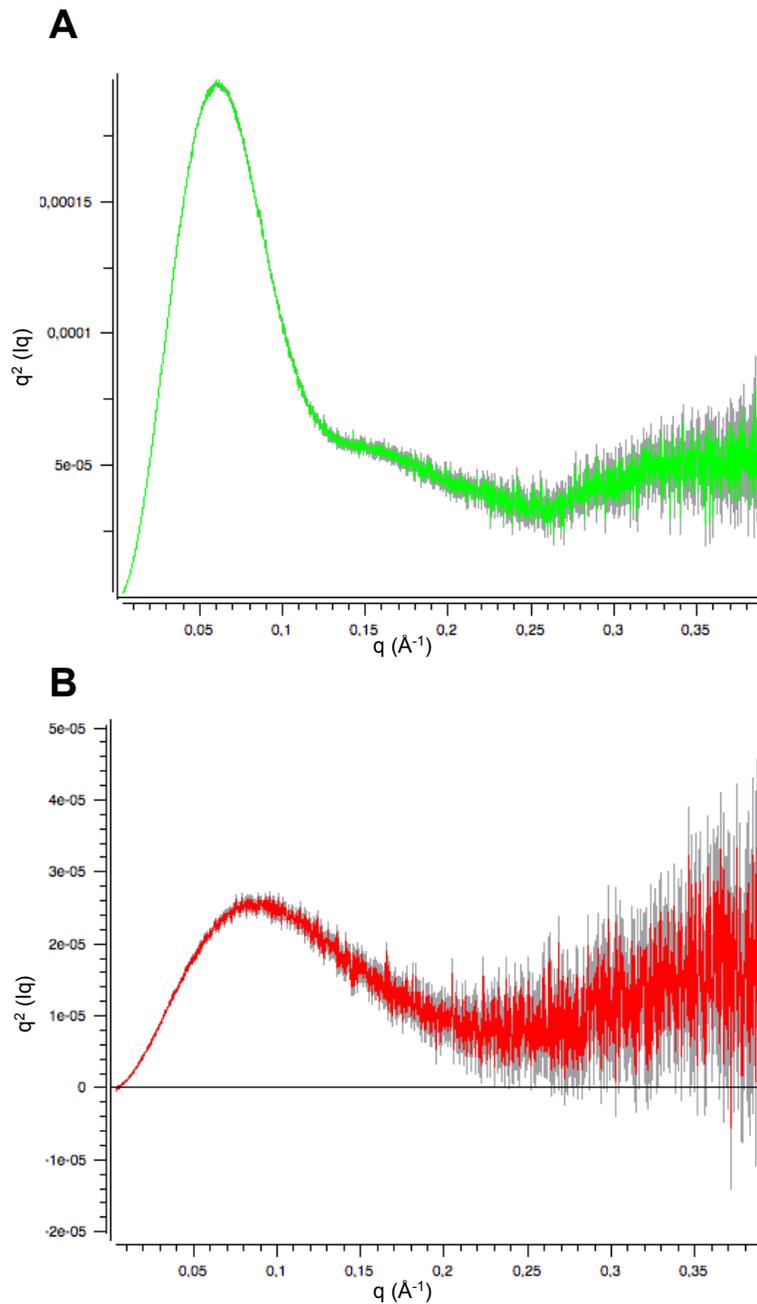


Figure 2.12 Kratky plots of (A) BSA control and (B) CHFR FHA scattering intensity data. The bell-shaped curves identifiable from both protein profiles is indicative of well-folded proteins.

However, the slight plateau exhibited within the higher q range (Putnam, 2007; Svergun *et al.*, 2013) is indicative that both proteins may be partially within an extended conformation in solution.

2.2.5 Model-based analysis of the CHFR FHA domain indicates the FHA domain in solution shares some characteristics with crystal structure counterpart

To obtain a low resolution, *ab initio* shape models of the CHFR FHA domain protein (13-180) in solution, DAMMIF (Franke and Svergun, 2009) was used to generate 20 different dummy atom models (Appendix 1, Section A1.6 Figure A1.23). Model alignment (DAMSUP) and averaging (DAMAV) was then performed to produce a single averaged model; generating a PDB model with most considerable shared bead position (or overlap) between all 20 models (Figure 2.13).

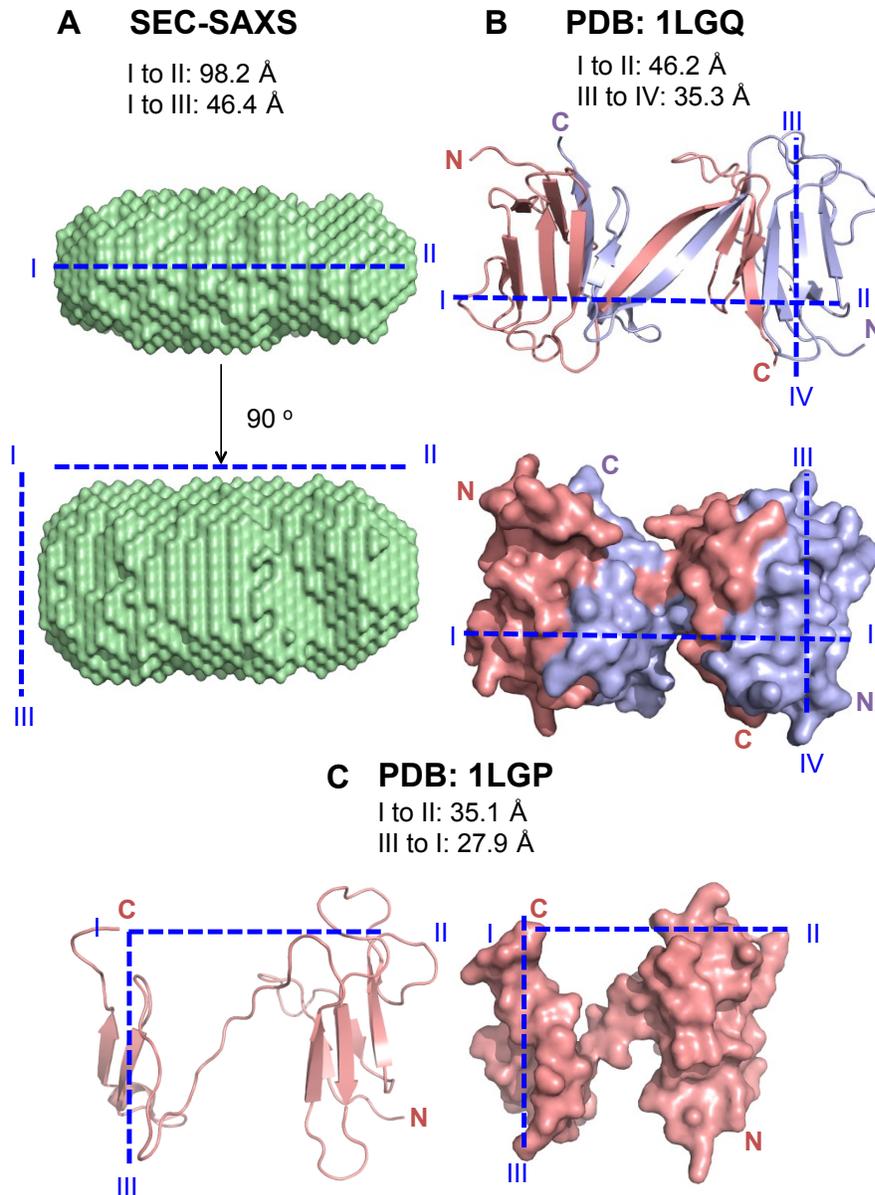


Figure 2.13 Model of the solution structure of the CHFR FHA domain (13-180) in comparison to experimental structures (Stavridi *et al.*, 2002). Space filled (surface) model of the CHFR FHA domain (13-180) obtained via SEC-SAXS. (B) The crystal structure of the CHFR FHA domain as a segment swapped dimer, encompassing residues 13 to 124 (PDB: 1LGQ); shown as both cartoon (top) and space filled model (bottom; Stavridi *et al.*, 2002). Individual FHA domain monomers are shown in pale blue and salmon. Measurement I corresponds to C46 at the C-terminus of one FHA protein molecule; whilst measurement II corresponds to E34 of a second protein. Measurement III to IV corresponds to residues K54 and K68 within the same FHA monomer, respectively. (C) The crystal structure of the CHFR FHA domain as a segment swapped dimer, encompassing residues 13 to 125 (PDB: 1LGP); shown as both cartoon (left) and space filled model (right; Stavridi *et al.*, 2002). Measurement I to II corresponds to the C-terminus of the FHA domain monomer at E125 to P51, within the same monomer. Measurement I to III corresponds to the FHA domain residue E125 to K92. Blue dashed lines indicate measured distances, using Pymol.

Unsurprisingly, measurements of the FHA domain width (vertical measurements) between FHA domain structures of residues 13 to 124 and 13 to 125 (PDBs 1LGQ and PDB 1LGP, respectively) are very similar (35.2 and 35.1 Å, respectively) (Figure 2.13, compare panels B and C). Differences in length (horizontal measurements) between both proteins is most likely resultant of the slight horizontal 'shift' and non-perfect overlap between individual FHA domain monomers (Figure 2.13, Panel B) within the segment swapped dimer (Stavridi *et al.*, 2002).

The height (vertical measurement) of the larger FHA domain protein in solution (encompassing CHFR residues 13 to 180; 46.4 Å) (Figure 2.13, Panel A) is strongly comparable to the observed length (horizontal measurement) of the segment swapped dimer (46.2 Å) (Figure 2.13, Panel B). This strongly suggests that the FHA domain dimer in solution has a comparable conformation, so some extent, to the crystal structure counterpart (PDB: 1LGQ, Stavridi *et al.*, 2002). The extended length (horizontal measurement) of the SAXS FHA domain model (Figure 2.13, Panel A, I to II) and extension to the space filled model (98.2 Å compared to crystal structure counterpart measurement of 35.3 Å) may be resultant of an extended loop region, with 56 extra residues within the N-terminus of the SEC-SAXS CHFR FHA domain protein.

While the low-resolution SAXS model does not provide sufficient data to specify secondary structural content or position, the extended and elongated conformation of the dimeric FHA domain is strongly comparable to the FHA domain experimental structure of the segment swapped dimer (Stavridi *et al.*, 2002). This dummy atom model suggests strong conservation of segment swapped dimerization in the solution, comparable to the X-ray experimental structure and AUC data (Stavridi *et al.*, 2002). In addition some conformational features within the segment swapped dimer X-ray experimental structure (PDB: 1LGQ) and the SEC-SAXS solution low-resolution model may be shared.

2.2.6 FL-CHFR increases the rate of K63-linked polyubiquitin chain formation in the presence of the Ubc13: Mms2 E2 heterodimer

Use of SEC with molecular weight standards (Figure 2.2 and Table 2.2) had identified Δ FHA1, -2, and -3 proteins as dimeric in solution. Kang et al (2002) has demonstrated that CHFR proteins without the CRD (Δ FHA- Δ CRD-CHFR) or without both the CRD and FHA domain (Δ FHA- Δ CRD-CHFR) are still active E3 ubiquitin ligases, with the capacity to form polyubiquitin chains *in vitro*. To verify that the selected FHA-deletion mutants used here, inclusive of the CRD (Δ FHA-CHFR, Figure 2.2, Panel A), are also active E3 ubiquitin ligases, ubiquitination assays using two independent E2 conjugating enzymes and assay (including UbcH5a or Ubc13: Mms2 heterodimer) were used.

Firstly, the requirement of individual reaction components within the ubiquitination assays for CHFR-specific polyubiquitination with either UbcH5a or Ubc13: Mms2 was assessed using a reaction component control (Figure 2.14).

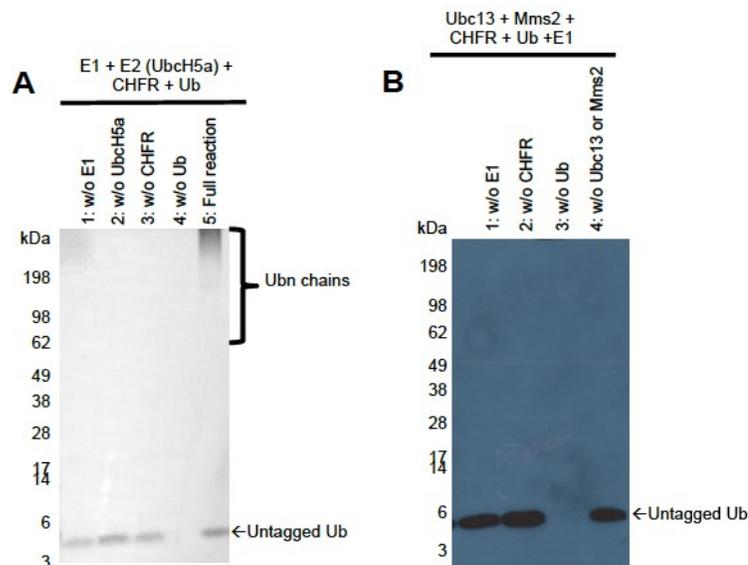


Figure 2.14 Validation of the CHFR ubiquitination assay by omitting reaction components. (A) To verify CHFR ubiquitin ligase assays, experiments were performed by omitting individual reaction components in the presence of either UbcH5a or (B) the Ubc13: Mms2 E2 complex. WO denotes ‘without’. Samples were resolved by SDS-PAGE, followed by western blot analysis using a primary antibody against ubiquitin (1: 500) to detect polyubiquitin chain formation.

By omitting any components within the assays, no polyubiquitin chains formation was observed and therefore demonstrating the assay dependence on each of these reaction components.

In addition, to assess the stability of the FL-CHFR protein within the ubiquitination and E2Scan assay buffers, protein samples were incubated at 37 °C and resolved by SDS-PAGE at different time points; corresponding to the maximum incubation periods for both ubiquitination and E2Scan assays (up to 3 hours for each, Figure 2.15).

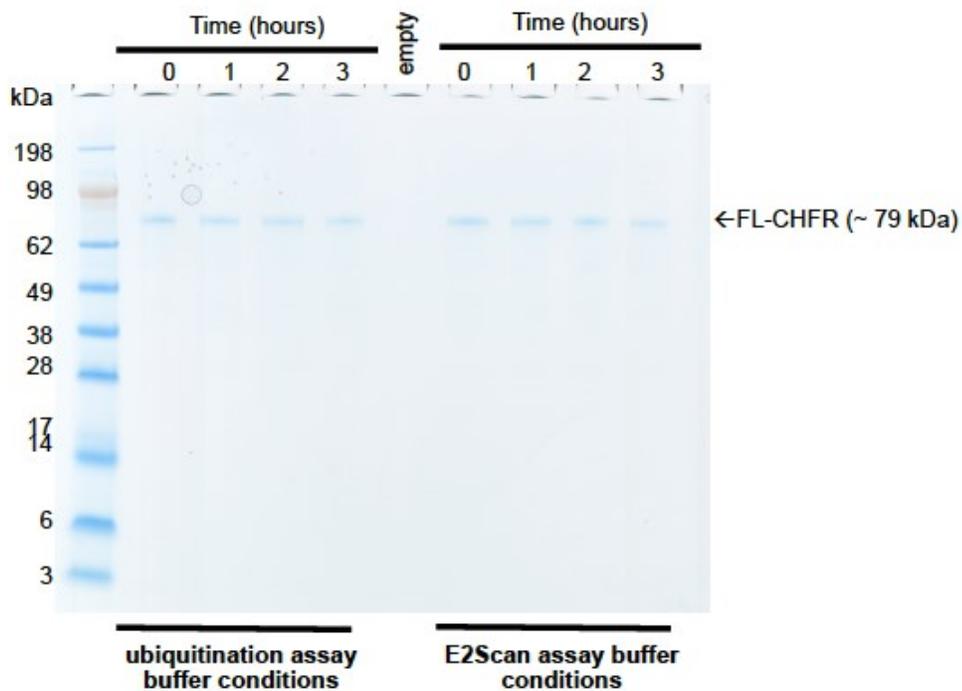


Figure 2.15 Verification of CHFR stability within buffer and temperature conditions of ubiquitination and E2Scan assay conditions. 1 μ M FL-CHFR was incubated in either ubiquitination (left) or E2Scan (right) buffer; with samples resolved prior to incubation (0 hours); and after 1, 2 or 3 hours incubation at 37 °C.

Across a 3- hour incubation at 37 °C, a single, clean band (at 79 kDa) corresponding to the FL-CHFR protein can be identified within the destained SDS-PAGE gel for both buffer conditions; indicative of CHFR's stability within both assay conditions.

Proceeding verification that CHFR ubiquitination assays require all individual reaction components for polyubiquitin chain formation (with either UbcH5a or Ubc13: Mms2 E2s) and CHFR's stability within the assay conditions; CHFR's activity as an E3 ligase was further characterised by performing the reactions in the absence and presence of the E3 protein within a time-series. In addition, it was assessed whether or not non-CHFR specific polyubiquitin chain formation occurs in the absence of the E3 ubiquitin ligase (Figure 2.16).

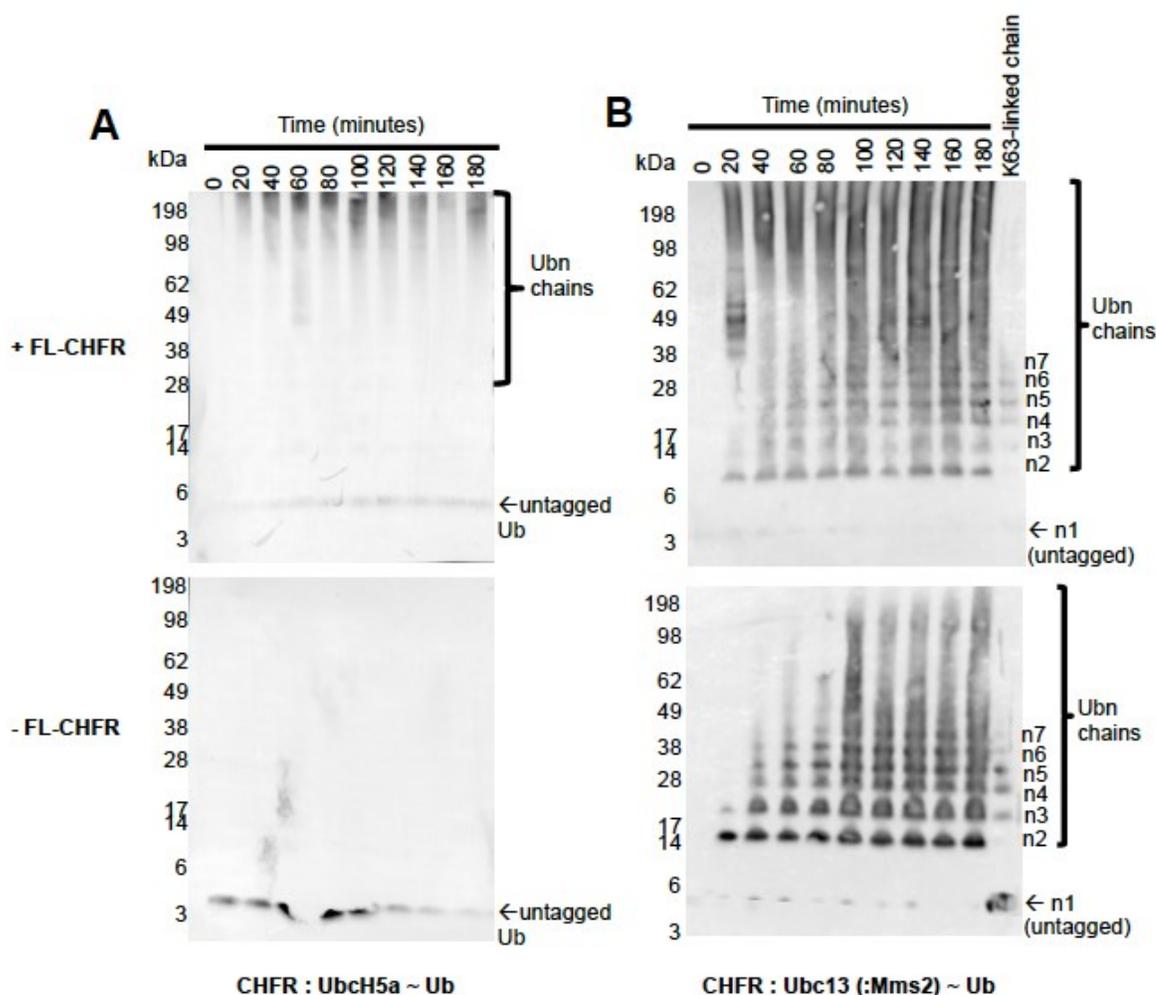


Figure 2.16 CHFR ubiquitination assays performed in the presence and absence of the E3 ubiquitin ligase with the E2s (A) UbcH5a or (B) Ubc13:Mms2. Samples were resolved immediately prior to incubation and the addition of ATP (0 minutes); or at 20 minute intervals following incubation at 37°C. For the Ubc13: Mms2 specific ubiquitination reaction (Panel B), a K63-linked chain was also included (right). Samples were resolved by SDS-PAGE, followed by western blot analysis using a primary antibody against ubiquitin (1: 500) to detect polyubiquitin chain formation.

In the presence of FL-CHFR, polyubiquitin chain formation is evident in the presence of UbcH5a after 20 minutes incubation at 37 °C (Figure 2.16, Panel A, top blot); with a similar amount of polyubiquitin chain formation evident after 180 minutes (3 hours). In contrast, no non-specific polyubiquitin chains are formed within the ubiquitination assay conditions in the absence of the FL-CHFR E3 ubiquitin ligase (Figure 2.16, Panel A, bottom blot), indicative that UbcH5a-specific polyubiquitin chain formation within the given assay conditions are specific to the E3 ubiquitin ligase activities of the FL-CHFR protein.

Similarly, FL-CHFR polyubiquitin chain formation also occurs after 20 minutes in the presence of the Ubc13: Mms2 heterodimeric E2 complex; whereby > n7 K63-linked polyubiquitin chains can be identified at 20 minutes; with an even greater quantity of ubiquitination product formed at 40 minutes and beyond (Figure 2.1, Panel B, top blot). However, in the absence of the FL-CHFR E3 ubiquitin ligase, at least n3 K63-linked polyubiquitin chains are present at 20 minutes; with a lower rate of ubiquitination product formed (in comparison to reaction conditions with FL-CHFR, Figure 2.1, Panel B, top blot) and amounts comparable to the CHFR counterpart reaction evident after 100 minutes.

Taken together, these results indicate that within the given ubiquitination assay reaction conditions; UbcH5a thioester conjugated polyubiquitin chains are dependent on the activity of FL-CHFR; with no chain formation evident in the absence of the E3 ubiquitin ligase. In contrast, the ubiquitination reaction conditions do facilitate the formation of (CHFR non-specific) K63-linked polyubiquitin chains via the Ubc13: Mms2 heterodimer. However, polyubiquitin chain formation occurs at a considerably lower rate in comparison to ubiquitination reactions within the presence of FL-CHFR. This is indicative that whilst chain formation can occur in the absence of the E3 ubiquitin ligase, an accelerated rate of CHFR-specific K63-linked polyubiquitin chain formation can be easily identified within the given assay conditions within at least the first 20 minutes of incubation at 37 °C.

2.2.7 Δ FHA CHFR dimers, inclusive of a CRD, are active E3 ubiquitin ligases

To assess whether the Δ FHA CHFR dimers, inclusive of the CRD, retain the capacity for formation of polyubiquitin chains in the presence of either UbcH5a or Ubc13: Mms2 E2 ubiquitin conjugating enzymes, a ubiquitination assay was performed using the Strep(II)-tagged CHFR Δ FHA2 (251-664) protein (Figure 2.17).

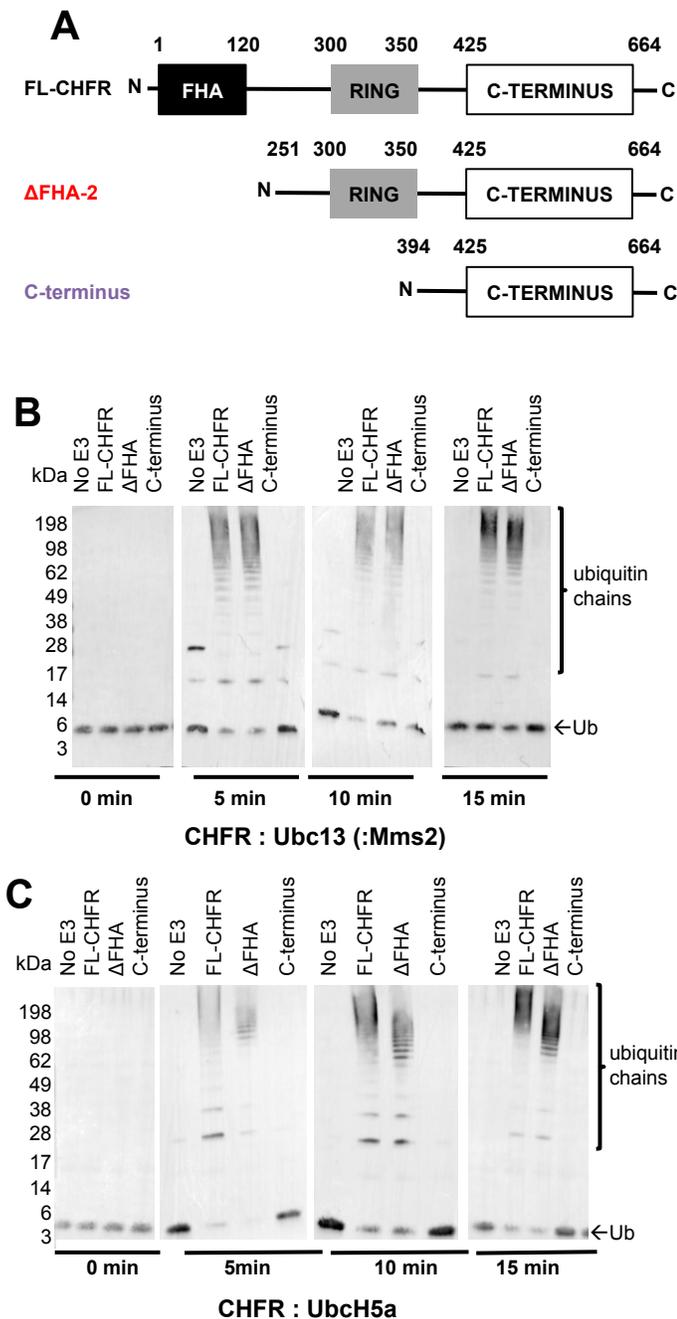


Figure 2.17 The Δ FHA CHFR dimers (with CRD) remain active as E3 ubiquitin ligases. (A) Ubiquitination assays were performed using the FL-CHFR protein (containing an FHA domain), CHFR Δ FHA-2 (without FHA domain) and the CHFR C-terminus (as a control for RING-domain specific E3 ubiquitin ligase activities). (B) Ubiquitination assays (also performed without CHFR, first lanes) were performed with Ubc13: Mms2 and (C) UbcH5a E2 ubiquitin conjugating enzymes. Samples were resolved by SDS-PAGE, followed by western blot analysis using a primary antibody against ubiquitin (1: 500) to detect polyubiquitin chain formation. No Ubc13: Mms2 or UbcH5a-mediated polyubiquitin chain formation is evident in both the presence of the CHFR C-terminal protein (without the CHFR RING-domain) and complete absence of any CHFR protein (Panel B and C respectively; final and first lanes of each blot per time point); indicative that all polyubiquitin chain formation identified is specifically produced via CHFR's RING domain E3 ubiquitin ligase activities and such chains are not a product of non-specific (non-CHFR mediated) polyubiquitin chain formation.

Within ubiquitination assays containing the Ubc13: Mms2 thioester conjugates, band intensities corresponding to amounts of polyubiquitin chain product are very similar 5, 10 and 15 minutes-post incubation (at 37 °C) (Figure 2.17, Panel B) in the presence of both FL-CHFR and CHFR Δ FHA-2. This is indicative that the CHFR FHA domain does not have an essential role in CHFR's activities as an E3 ubiquitin ligase and producing K63-linked polyubiquitin chains, corroborating with the findings of Kang *et al* (2002). Most importantly, this is first demonstration that the CHFR RING E3 remains active without the FHA domain, but with the CRD still retained (Figure 2.17, Panel A).

However, within ubiquitination assay conditions containing the UbcH5a E2 enzyme (Figure 2.17, Panel C), efficiency of polyubiquitin chain formation may be lower in the presence of the CHFR Δ FHA-2 protein, in comparison to full-length counterparts. At 5, 10 and 15 minutes post-incubation, the molecular weight of polyubiquitin chains formed in the presence of the Δ FHA-2 protein do not exceed 198 kDa; in comparison to FL-CHFR reaction counterparts. This may indicate a role for the CHFR FHA domain in optimizing very high molecular weight UbcH5a-specific polyubiquitin chain formation and/or maximizing efficiency of chain formation.

2.2.8 Identification of new E2 ubiquitin conjugating enzymes responsible for CHFR ubiquitination

In order to identify new E2 substrates for FL-CHFR-mediated polyubiquitin chain formation, an E2 Scan assay, consisting of 34 different E2 ubiquitin conjugating enzymes, was to screen for new thioester conjugate enzymes (Figure 2.18).

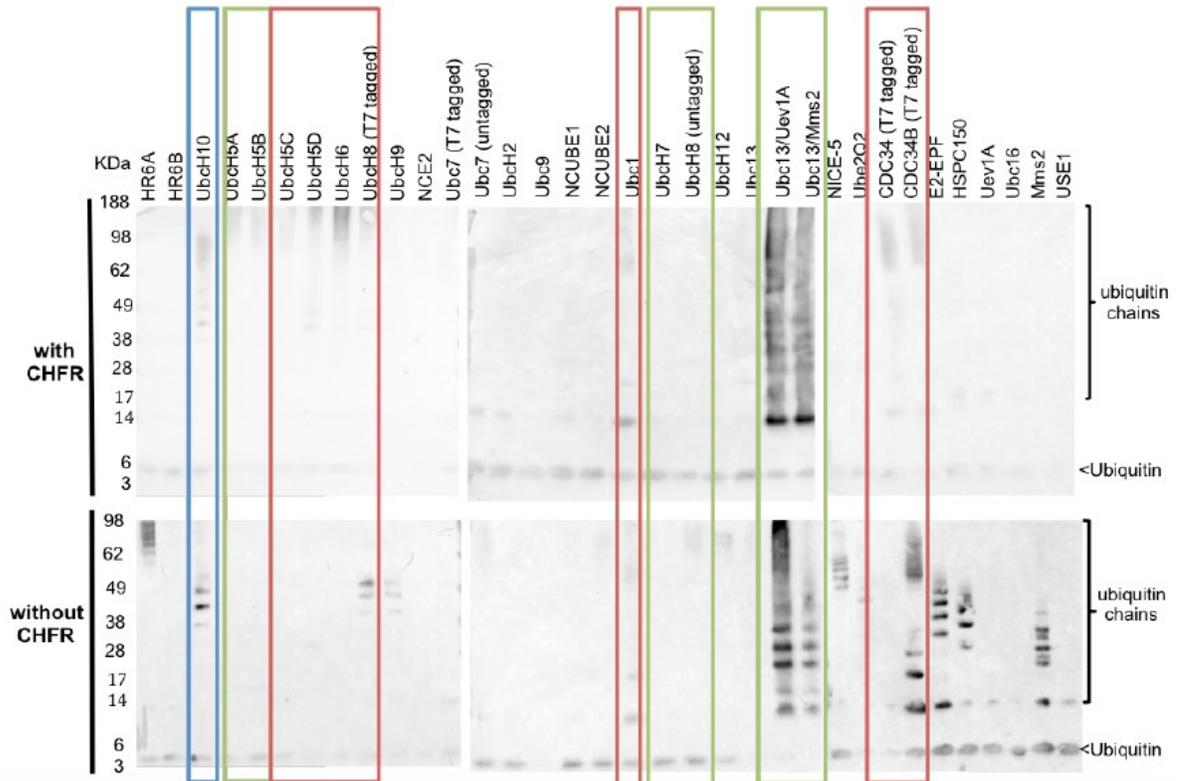


Figure 2.18 Identification of new E2 ubiquitin conjugating enzymes responsible for CHFR ubiquitination using an E2Scan ubiquitination assay. A ubiquitination assay was performed using 34 different E2 ubiquitin conjugation enzymes (Appendix 2.1, Table A2.3) in the presence (top 3 panels) and absence (lower 3 panels) of the FL-CHFR E3 ubiquitin ligase. Samples were resolved by SDS-PAGE, followed by western blot analysis using a primary antibody against ubiquitin (1: 500) to detect polyubiquitin chain formation. Green boxes indicate previously verified E2s responsible for CHFR ubiquitination, whilst red boxes highlight newly identified E2 substrates. With comparable rates of CHFR ubiquitination observed in the presence of the Ubch10 E2 enzyme, a blue box is included.

A considerably high amount of polyubiquitin chain products in both the presence and absence of FL-CHFR strongly correlates with non-CHFR specific K63-linked polyubiquitin chain formation previously identified (Figure 2.18, Panel B, lower blot). In addition, Anon (2013) indicate that the ubiquitin loading activity of Ubc13, Ubc13: Uev1A and Ubc13: Mms2 is 70 %, as indicated by the high amount of polyubiquitin chain formation specifically present within assay conditions containing the Ubc13 heterodimeric complexes. Notably, whilst the ubiquitin loading activity of Ubc13 is 70%, the absence of the Mms2 and Uev1A pseudo E2 enzymes results in no CHFR specific (Figure 2.18, upper panel) and no CHFR non-specific (Figure 2.18, lower panel) polyubiquitin chain formation.

CHFR polyubiquitin chain formation in the presence of Ubch5a (Bothos *et al.*, 2003, Chatrvedi *et al.*, 2002); in addition to Ubch5B, Ubch7, Ubch8 (untagged), Ubch13:Uev1A and Ubch13:Mms2 (Figure 2.18, green boxes; Bothos *et al.*, 2003) have been previously reported.

New E2 ubiquitin conjugating enzymes identified as responsible for CHFR E3-specific polyubiquitin chain formation (Figure 2.18, red boxes) include: Ubch5C, Ubch5D, Ubch6, Ubc1, Ubch8a and CDC34; with some (non-E3) ubiquitin chain formation present in ubiquitination assays containing Ubch8 (T7 tagged) and CDC34B but without FL-CHFR. The noticeably lower blot signal intensity within lanes containing assay conditions inclusive of these E2 ubiquitin-conjugating enzymes, in comparison to Ubch13-heterodimeric counterparts, is most likely to be attributable to the lower percentage of E2 loading activity; calculated at 50%, 60-70%, 50% and 50% for all Ubch5a isoforms, Ubch6, Ubc1 and CDC34/34B, respectively (Anon, 2013).

Whilst Chatrvedi *et al* (2002) have reported Ubch10 as an E2 ubiquitin conjugating enzyme responsible for CHFR-mediated polyubiquitin chain formation, the strongly comparable amounts of polyubiquitin chains formed in both presence and absence of FL-CHFR (Figure 2.18, blue box) make it very difficult to differentiate as to whether polyubiquitin chains formed in the absence of CHFR after 3 hours are specific to CHFR's activity as an E3 ubiquitin ligase or simply non-E3 dependent chain formation.

Stability of the FL-CHFR protein was previously verified via SDS-PAGE analysis of the incubated E3 ubiquitin ligase in identical assay conditions (Figure 2.15).

Assay conditions within the E2Scan assay were also confirmed as optimal for CHIP E3 ubiquitin ligase formation of polyubiquitin chains (in the presence of the Ube2D4 E2 ubiquitin conjugating enzyme) and CHFR polyubiquitin chain formation verified as E2-dependent (Figure 2.19).

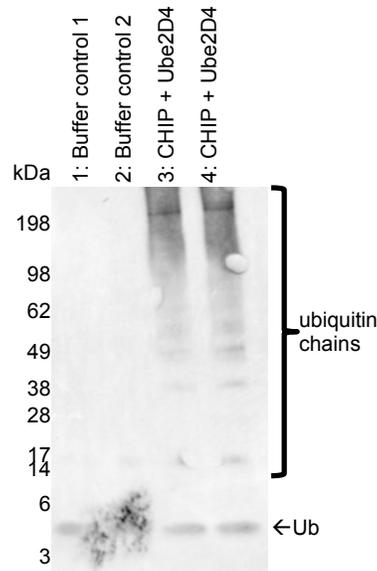


Figure 2.19 CHFR polyubiquitin chain formation is E2 dependent, with E2Scan assay conditions suitable for CHIP E3-mediated polyubiquitin chain formation. Buffer controls (Lanes 1 and 2) included the entire ubiquitination reaction composition, with the exception of any E2 ubiquitin conjugating enzymes. The lack of polyubiquitin chain formation identified verifies that CHFR-specific polyubiquitin chain formation within the E2Scan assay is E2 dependent. Reaction mix composition and experimental conditions were also deemed optimal for CHIP E3-specific formation of polyubiquitin chains (Lanes 3 and 4); with ubiquitination products produced in conditions identical to those used in the CHFR-specific E2Scan assay. Samples were resolved by SDS-PAGE, followed by western blot analysis using a primary antibody against ubiquitin (1: 500) to detect polyubiquitin chain formation.

2.2.9 CHFR-specific polyubiquitin chain formation in the presence of the CDC34 and CDC34B E2 enzymes may be T7 tag dependent

Noticeable, within the E2Scan assay, FL-CHFR specific polyubiquitin chain formation appears to be evident in the presence of the T7-tagged CDC34 E2 ubiquitin conjugating enzyme (Figure 2.18, comparing top and bottom panels). In addition, T7-tagged CDC34B ubiquitin conjugation in both the presence and absence of FL-CHFR was also observed, potentially indicating CDC34 and/or CDC34B as E2 ubiquitin conjugating enzymes responsible for CHFR-specific polyubiquitin chain formation.

In an attempt to reproduce the E2Scan assay results and to rule-out any polyubiquitin chain formation that is E2 tag-specific, a ubiquitination assay was performed using identical assay conditions, but untagged CDC34 and untagged CDC34B proteins (Figure 2.20).

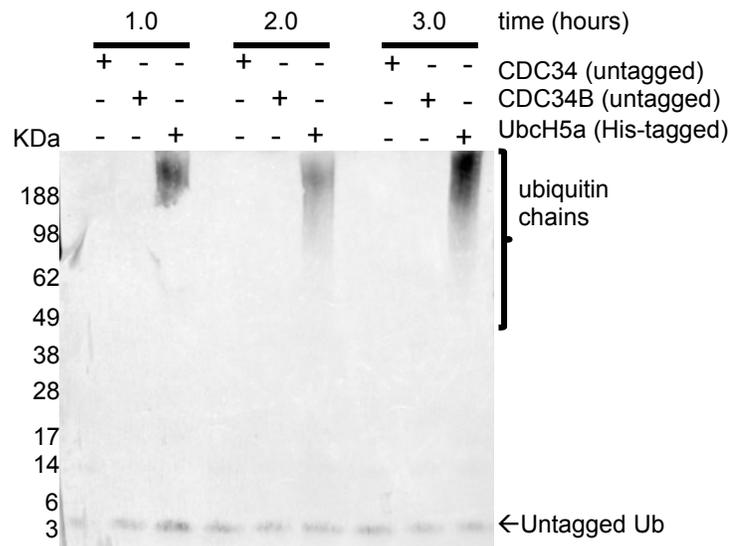


Figure 2.20 CHFR polyubiquitin chain formation in the presence of the CDC34 and CDC34B E2 enzymes is T7-tag dependent. Ubiquitination assays were performed using untagged CDC34, untagged CDC34B (both Ubiquigent) or His-tagged Ubch5a in the presence of FL-CHFR; with reactions stopped at 1, 2 or 3 hours post-incubation at 37 °C. Samples were resolved by SDS-PAGE, followed by western blot analysis using a primary antibody against ubiquitin (1: 500) to detect polyubiquitin chain formation.

Whilst polyubiquitin chain formation within identical assay conditions (inclusive of E2 storage buffer) is evident for the recombinantly expressed His5-tagged Ubch5a E2 enzyme and FL-CHFR E3 ubiquitin ligase across all three time points, no chain formation is evident in the (untagged) CDC34 and CDC34B E2 containing counterparts.

This is indicative that any observed polyubiquitin chain formation observed within the E2Scan assay and in the presence of the FL-CHFR protein (Figure 2.18, top panel) is most likely to be resultant of CDC34 and CDC34B T7-tag- specific chain formation (Figure 2.20).

2.2.10 Phylogenetic analysis of E2 ubiquitin conjugating enzymes exhibits some shared evolutionary relationship between E2s responsible for CHFR-mediated polyubiquitin chain formation

Whilst the E2Scan ubiquitination assay provided insight into the specific E2 ubiquitin conjugating enzymes recognized by the CHFR proteins and responsible for CHFR-mediated polyubiquitin chain formation *in vitro*, the relationship between biological similarity (or dissimilarity) concerning the sequences of E2s and presence of polyubiquitin chain formation remains unexplored.

To determine any biological relationship between the E2 enzymes within the E2Scan kit used to explore CHFR-mediated polyubiquitin chain formation (Figure 2.18) a multiple sequence alignment was initially constructed. Bayesian inference (derived from a Markov chain Monte Carlo (MCMC) algorithm) and Maximum Likelihood Estimations (MLEs) were used to generate two independent phylogenetic trees (Figure 2.21).

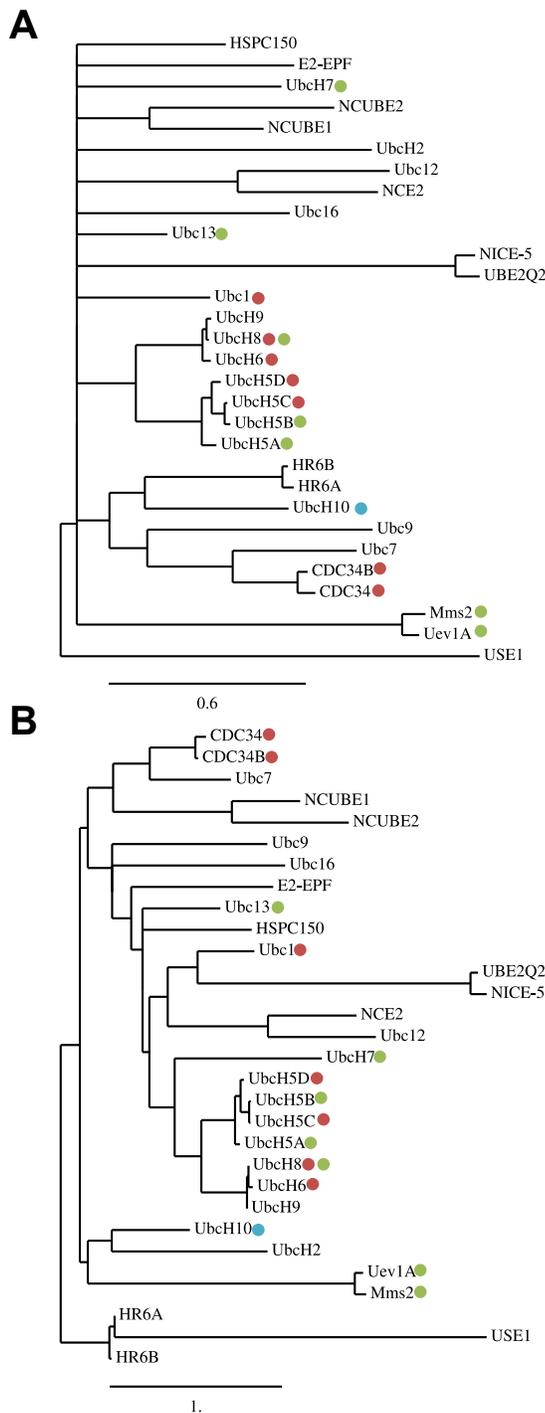


Figure 2.21 Unrooted, phylogenetic trees of E2 enzymes within the E2Scan ubiquitination assay kit. Unrooted, phylogenetic trees were constructed using (A) Bayesian inference with MCMC (using MrBayes; Huelsenbeck and Ronquist, 2001) and (B) maximum likelihood cladistics (using PHYLIP; Retief, 2000); inferred from sequence alignment containing 30 different E2 enzymes using PHYLIP (Version 3.695; Retief, 2000). The Jones-Taylor-Thornton (JTT; Jones *et al.*, 1992) probability model for rate of amino acid change was used to calculate maximum likelihood cladistics. Scale bars for edge lengths are also indicated. Green circles indicate previously verified E2s responsible for CHFR-mediated polyubiquitin chain formation, whilst red circles highlight newly identified E2 substrates (via E2Scan assay, Figure 2.18). With comparable rates of CHFR-mediated polyubiquitin chain formation observed in the presence of the UbcH10 E2 enzyme, a blue box is also included.

Phylogenetic analysis of the 30 E2 ubiquitin conjugating enzyme sequences using both statistical methods revealed strong sequence similarity between the four Ubch5 isotypes (A, B, C and D); as exhibited by the short edge lengths between connecting nodes (corresponding to a short amount of evolutionary time) and distinct clade (Figure 2.21). This strongly corroborates with the polyubiquitin chain formation observed *in vitro* across all 4 ubiquitination assays containing the individual Ubch5 isoforms (Figure 2.18).

Whilst an incredibly short amount of evolutionary time (and therefore high sequence similarity) is evident between the Ubch8, Ubch6 and Ubch9 E2 ubiquitin enzymes (using both statistical methods, Figure 2.21, Panels A and B) only Ubch8 and Ubch6 were responsible for CHFR-mediated polyubiquitin chain formation *in vitro* (Figure 2.18).

The Ubc1 E2 ubiquitin conjugating enzyme was also identified as responsible for CHFR-mediated polyubiquitin chain formation *in vitro*, with an independent node indicative of evolutionary dissimilarity between other E2s using Bayesian Inference (Figure 2.21, Panel A). However, a common node shared with Ubc1, UBE2Q2 and NICE-5 via MLEs (Figure 2.21, Panel B) may indicate some sequence similarity; not attributed to lack of polyubiquitin chain formation upon CHFR in the presence of UBE2Q2 and NICE-5 (Figure 2.18).

A similar ambiguity in associated clade and observed ubiquitination activity is exhibited for the Ubch7 E2 enzyme. A shared internal node (and therefore biological relationship) is inferred between Ubch7 and other E2s identified as responsible for CHFR-mediated polyubiquitin chain formation (such as the Ubch5a isoforms, Ubch8 and Ubch6) via tree construction using MLE (Figure 2.21, Panel B); but a unique node (and therefore sequence evolution) inferred via Bayesian statistical analysis (Figure 2.21, Panel A).

Unsurprisingly, a consensus between Bayesian inference and MLE statistical methods used to construct a unique clade encompassing both pseudo E2 enzymes Uev1A and Mms2 (Figure 2.21, Panel A and B, respectively) may be reflective of their high sequence similarity and their capacity to mediate CHFR-specific polyubiquitin chain formation in the presence of Ubc13 (Figure 2.18).

2.3 DISCUSSION

2.3.1 A model of the dimeric CHFR protein

X-ray structures of individual CHFR FHA (PDB: 1LGQ, Stavridi *et al.*, 2002) and C-terminal (PDB: 2XP0, Oberoi *et al.*, 2010) domains indicate proteins may be dimeric in solution. Conformation of segment-swapped specific dimerization was unachievable using X-ray crystallography structures and AUC alone (Stavridi *et al.*, 2002), with crystal contacts within the CHFR C-terminus crystal structure (PDB: 2XP0) potentially indicative of a dimeric protein in solution (Oberoi *et al.*, 2010). In addition, the biophysical characteristics of the FL-CHFR protein are incredibly poorly understood, with potential domain-specific structural roles or oligomeric roles in mediating CHFR specific polyubiquitin chain formation currently unknown.

FHA domain (13-180) dimerization in solution was verified by SEC-SAXS experiments with both the calculated molecular weight (Table 2.2) and measured dimensions (Figure 2.13) corroborating with the segment swapped dimer crystal structure (PDB: 1LGQ, Stavridi *et al.*, 2002). However, Δ -FHA CHFR proteins retain the capacity to form polyubiquitin chains *in vitro*, inclusive of the CRD (Figure 2.17), suggesting dimerization of the FL-CHFR protein is most likely mediated by the CRD.

Calculation of FL-CHFR and Δ -FHA protein molecular weights (and deduction of respective oligomeric states) using SEC and MW standards indicate both proteins are dimeric in solution (Figure 2.2, Panel B; Table 2.2), with FL-CHFR dimerization

also verified by native PAGE (Figure 2.4). In addition, dimerization of the CRD (Figure 2.2, Panel B; Table 2.2) and higher order oligomerization (Figure 2.3) suggests FHA-independent dimerization of the FL-CHFR protein, potentially via the CRD. However, the percentage differences in molecular weights calculated via SEC with MW standards compared to those of the actual amino acid sequences using ProtParam (Table 2.2) may be resultant of poor resolution, and therefore elution volumes, between protein peaks and their preceding protein aggregate/impurity counterparts (Figure 2.2, Panel B). Alternative techniques, such as SEC-MALLS (using a pre-optimized analytical SEC column) or native mass spectrometry may provide a more accurate size determination of the different CHFR proteins.

Whilst the CHFR orthologue Dma1 is an obligatory homodimer (Johnson *et al.*, 2012), it is unclear whether CHFR dimerization is essential for polyubiquitin chain formation. Previous studies have shown that the central RING domain alone (Δ FHA- Δ CRD-CHFR, residues 143 to 360 and 267 to 360) without the CRD or FHA domains is capable of polyubiquitin chain formation, with higher rates of polyubiquitin chain formation were also observed comparison to a CRD-deletion counterpart (inclusive of the FHA domain, Δ CRD-CHFR, residues 1 to 360) (Kang *et al.*, 2002). In addition, an *in vitro* role for the FHA or CRD in mediating CHFR-mediated polyubiquitin chain formation was not identified by Kang *et al.* (2002). Scolnick and Halazonetis (2002) have demonstrated an importance for both the CHFR FHA and CRD in maintaining the antephasic checkpoint. Using a SAOS2 cell line transfected with a Δ FHA- Δ CRD-CHFR domain-double mutant, endogenous CHFR (FL-CHFR) lost its capacity to delay mitotic progression in the presence of mitotic stress (Scolnick and Halazonetis, 2002). Taken together, this suggests that *in vitro*, the smaller Δ FHA- Δ CRD-CHFR domain-double mutant may have a capacity to form polyubiquitin chains at a higher rate (Kang *et al.*, 2002) as a (monomeric) RING E3 ubiquitin ligase, in comparison to FHA-domain (dimerizing) counterparts (Kang *et al.*, 2002). However, the important role of the FHA domain and/or C-terminal CRD in maintaining the antephasic checkpoint *in*

vivo (Scolnick and Halazonetis, 2002) suggests that FL-CHFR may be an obligate dimer in order to facilitate its interactions with other antepause checkpoint proteins, independent of the ubiquitination enzymatic cascade.

The FL-CHFR protein was also assessed for CC content, to rule out any additional structural features independent of the FHA domain and CRD that may be responsible for its dimerization in solution. Using of two independent bioinformatical coiled-coil prediction servers (COILS and PairCoil2, Figures 2.5 and 2.6 respectively), the absence of heptad repeats within the amino acid sequence of the entire protein indicates CHFR dimerization in solution is not mediated by coiled-coils. In contrast, the RING E3 ubiquitin ligase CHIP homodimerizes at each promoter; whereby N-terminal hairpins aggregated to establish a four-helical bundle (Zhang *et al.*, 2005). Negative staining of the full- with a single GFP-tag at the protein N or C-terminus of the protein (Corbalan *et al.*, 2013) would be indicative of domain specific dimerization within the FL-CHFR protein.

Diverse ranges of strategies are enlisted by E3 ubiquitin ligases to mediate oligomerization. Extensive coiled-coil formation within the RNF8 RING E3 homologue (Campbell *et al.*, 2012) is essential in its activities targeting substrates for K63-linked polyubiquitin chain formation in the presence of the Ubc13:Mms2 heterodimeric E2 complex. Whilst CHFR lacks coiled-coil domains, the segment-swapped dimer formation within the FHA domain crystal structure (Stavridi *et al.*, 2002) may compensate for this. However, it remains unclear as to whether CHFR RING (or otherwise) domain multimer formation is essential for optimal catalytic activity as a E3 ubiquitin ligase, particularly considering the *in vivo* findings concerning CHFR domain deletion mutants and maintenance of the antepause checkpoint, observed by Scolnick and Halazonetis (2002).

2.3.2 Identification of new E2s responsible for CHFR-mediated polyubiquitin chain formation

In order to identify E2s responsible for CHFR-mediated polyubiquitin chain formation *in vitro*, 34 different enzymes were assessed within ubiquitination assays for their capacity to mediate polyubiquitin chain formation, specifically in the presence of the E3 to rule out non-specific chain formation (Figure 2.18). E2 ubiquitin conjugating enzymes newly identified as responsible for polyubiquitin chain formation via the CHFR RING E3 include: UBCH5C, UbcH5D, UbcH6, UbcH8 and Ubc1 (Figure 2.18). Phylogenetic analysis of E2s using Bayesian inference and maximum likelihood (MLE) bioinformatics techniques identified a close evolutionary relationship between all four UbcH5A isoforms (A to D), indicative of their sequence similarity (Figure 2.21). In addition, the shared internal clade evident for UbcH8 and UbcH6 also indicates a close evolutionary relationship between selected E2s responsible for CHFR-mediated polyubiquitin chain formation, corroborating with polyubiquitin chain formation observed within *in vitro* ubiquitination assays (Figure 2.18). Independent phylogenetic analysis of E2s conducted by Sheng *et al* (2012) also identified UbcH5A, B and C isoforms within a distinct clade; comparable to a shared internal node also evident between UbcH6, UbcH8 and UbcH9 counterparts. The phylogenetic analysis conducted by Sheng *et al* (2012) utilized sequences specifically corresponding to the strongly conserved E2 ubiquitin conjugating enzyme UBC folds (Figure 1.3); suggesting CHFR recognition of E2 enzymes and resulting E3 ubiquitination is most likely resultant of a specific recognition of UBCs within closely evolutionary related E2s.

CHAPTER 3: Homology modelling of the CHFR RING domain

INTRODUCTION

Homology modelling

Current understanding of protein 3D structures has been significantly accelerated by advancements in X-ray crystallography, NMR and 3D-EM. As of March 2016, 108,481 protein structures are deposited within the PDB (Protein Data Bank, 2015; Abola *et al.*, 1987; Berman 2000; Rose, 2015). Considering the current UniProtKB/TrEMBL (2016) protein database release contains over 60 million sequence entries, the acquirement rate of DNA sequences is likely to surpass attainment of protein structures (Drozdetskiy *et al.*, 2015).

In order to gain a greater understanding of CHFR protein structure and its interactions with key ubiquitination proteins (such as E2 ubiquitin conjugating enzymes and ubiquitin); preceding CHFR-mediated polyubiquitin chain formation of the protein and potential down-stream ubiquitination of protein kinase substrates, structural information regarding the CHFR central RING domain is essential.

However, previous attempts to express and purify the CHFR RING domain (alone) proved unsuccessful. In addition, attempts to crystalize the full-length CHFR protein (Isoform 2, Accession: NP_001154817.1) within a wide range of crystallography conditions did not produce the protein crystals required for X-ray diffraction experiments. As of September 2017, no experimental structure (obtained via X-ray crystallography or NMR) nor model of the CHFR central RING domain is available.

Where experimental structures are unavailable, homology (or comparative) modelling of protein structures involves prediction of a protein 3D structure based upon an alignment between the protein sequence of interest (the target) and one or more related sequences (template); whereby the 3D structure is already known (Eswarn *et al.*, 2006, Ginallski *et al.*, 2008). Therefore, structural similarity between

template and target proteins is assumed, since sequence similarity is detectable (Marti-Renom *et al.*, 2000). The process of homology modelling can be categorized into 5 discrete steps: searching for templates, template selection, production of a target-template alignment, model production and evaluation (Figure 3.1, Sanchez and Sali, 2000).

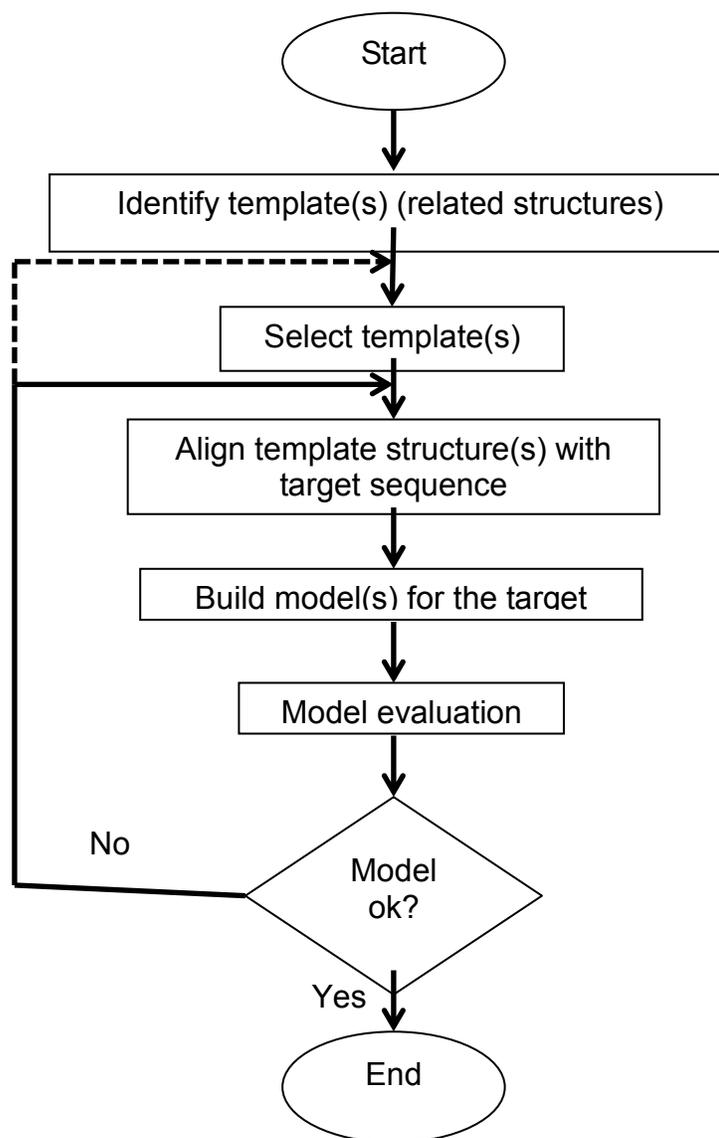


Figure 3.1 Overview of the homology modelling process (Adapted from Sanchez and Sali, 2000).

Multiple template (known) structures are obtained from the Protein Data Bank (PDB) (Abola *et al.*, 1987) through BLAST searches (Altschul *et al.*, 1997) performed using the target sequence (Sanchez and Sali, 2000). Appropriate

templates are selected on the basis of highest percentage of sequence identity between the target and template and fewest gaps within the alignment (Sanchez and Sali, 2000). Blundell and Hubbard (1987) previously demonstrated that difference in Root Mean Square Deviation (RMSD) between different proteins with detectable sequence identity (based on their respective coordinates) is related to resolution of the individual structures. High resolution and resolution of crystallographic-derived datasets are therefore desirable in obtaining high-quality models, since the sequence alignment and coordinates of the template structure are used within the modelling process (Cong *et al.*, 2010). Alternatively, the number of restraints per residue taken into consideration when selecting NMR-sourced templates, for the same reason (Sanchez and Sali, 2000). Structural equivalence between the selected target(s) and templates is obtainable by producing an alignment, ideally with at least 40% identity between target and template (Venclovas *et al.*, 1997). Tramontano *et al.* (1998) estimated models produced from a target: template identity of ~30% still produces acceptable models. However, target and template sequence identity less than 50% is most likely to result in RMSD between experimentally determined structures and respective model to exceed 1Å (Hubbard and Blundell, 1987) and production of more unreliable models (Eloffsson *et al.*, 2002). Generally, models exhibiting 30-50% sequence identity to their targets are likely to share ~80% of their structure (Ginalski *et al.*, 2008). Models are then calculated on the basis of both alignment and template coordinates, with the majority of homology modelling implemented using either fully automated or semi-automated protocols. The evaluation stage utilizes both internal evaluation resources, specific to the programme used to calculate the model itself; and external evaluation tools that are independent of this (Sippl, 1993; Lüthy *et al.*, 1992). In addition, energy minimization experiments may also be used to further refine models produced (Venclovas *et al.*, 1997).

A wide range of web-based applications and stand-alone software packages are available to facilitate production of 3D models by homology modelling methods. MODELLER is a homology modelling software package, which facilitates the

production of 3D models by the satisfaction of spatial restraints (Sali and Blundell, 1993; Eswarn *et al.*, 2006). Developed by Kaufmann *et al* (2010), Rosetta is an integrated software package utilized for homology modelling and protein 3D structure prediction by exploring conformational space and guiding structural evaluation by template energetic profiles and Monte Carlo sampling approaches (Rohl *et al.*, 2004). The SWISS-MODEL server provides a range of computational tools to facilitate homology modelling; including the front-end graphical SWISS-PDB Viewer, SWISS-MODEL workspace (Arnold *et al.*, 2006; Guex and Peitsch, 1997) to conduct homology modelling experiments (Guer and Peitsch, 1997) and SWISS-MODEL repository (Koppe and Schwede, 2004).

Whilst a large number of homology modelling resources are fully automated, manual intervention remains essential in ensuring the most accurate models are produced containing minimal errors (Eswarn *et al.*, 2006; Venclovas *et al.*, 1997).

By modelling the structure of the CHFR central RING domain, the biochemical basis of E2 and ubiquitin recognition, preceding CHFR mediated polyubiquitin chain formation or protein substrate ubiquitination, can then be modelled.

Molecular Dynamic (MD) simulations

Whilst the structural studies of static molecules (for example, via attainment of X-ray crystallographic data and therefore protein structures) provide important information underpinning protein function, the complex array of activities within dynamic biological systems remains insufficiently understood (Schlick, 2010; Hess *et al.*, 2008). The interdisciplinary field of molecular dynamics (MDs) and simulations generated thereof encompass computational approaches to investigating statistical mechanical properties of atoms and molecules, monitoring the dynamics of molecules (in space and time) whilst exploring important dynamic and structural properties (Schlick, 2010).

The most commonly used water models within MD simulations are rigid and non-polarizable; featuring differences in target properties (in comparison to real water), bond geometry (regarding H-O-H bond angle and lengths) and charge distribution present (Vega *et al.*, 2009). Specific geometric simulation restraints within MD simulations are typically atomic or course grained, generated by more than one simulation and includes a solvent (Abraham *et al.*, 2015; Pall *et al.*, 2015; Tsui and Case, 2001).

Computational resources available for MD simulations include GROMACS (Van der Spoel *et al.*, 2005), CHARM (Brooks *et al.*, 1983) and within the AMBER suite of programmes (Case *et al.*, 2005).

Incorporation of metal ion centres in Molecular Dynamic (MD) simulations

The interdisciplinary field of molecular dynamics (MDs) and simulations generated thereof encompass computational approaches to investigating statistical mechanical properties of atoms and molecules, monitoring the dynamics of molecules (in space and time) whilst exploring important dynamic and structural properties (Schlink, 2010). However, the incorporation of metal ions within molecular dynamic force field parameters is required prior to implementation of molecular dynamic (MD) simulations (Peters *et al.*, 2010). To address these issues, Peters *et al* (2010) developed the Metal Centre Parameter Builder (MCPB) programme; facilitating the inclusion of metal ion centres in MD studies of metalloproteins.

As of February 2016, more than 43,000 PDB structures include metal ions; with the majority including zinc (Zn), magnesium (Mg), calcium (Ca), iron (Fe), and sodium (Na) metal ions (MetalPDB 2016, Andrenini *et al.*, 2012). Considering the critical roles of metal ions underpinning protein stability, function and structure; MD simulations modelling metalloproteins offer an important insight into protein dynamic and structural protein features (Peters *et al.*, 2010).

Three main strategies utilized in the incorporation of metal ions within MD force fields include non-bonded, bonded plus electrostatic and cationic dummy models (Figure 3.2) (Peters *et al.*, 2010).

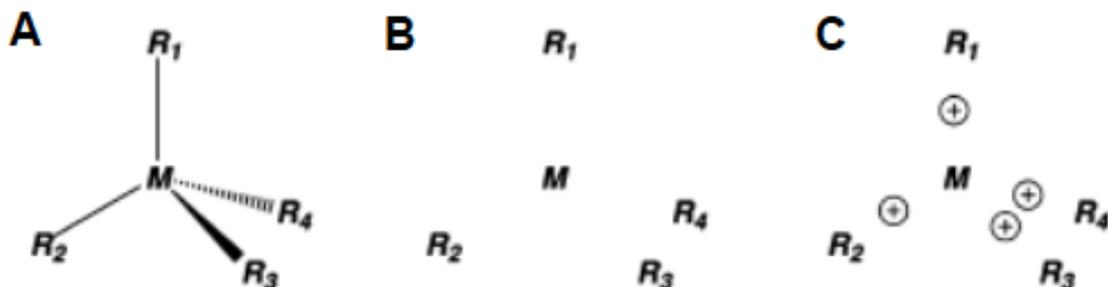


Figure 3.2 Three independent strategies for inclusion of metal atoms into molecular mechanics force fields (adapted from Peters *et al.*, 2010). (A) The bonded model defines various parameters between both metal atom and ligand; including dihedrals, bonds and angles (Peters *et al.*, 2010). However, the fixed coordinate number within bonded models does not facilitate modeling changes in coordinate or ligand switching (Li *et al.*, 2013). (B) Non-bonded models exclude these parameters and instead model such integrations exclusively via Van der Waals and electrostatics (Peters *et al.*, 2010). Metal ion coordination is more flexible in comparison to non-bonded model counterparts, whereby both ligand switching and coordinate number are permitted (Li *et al.*, 2013). However, an oversimplification of the ligand (coordinating residue):l on interaction can be observed; with the single point-charge assigned to metal ions is a poor representation of ion charge distribution, considering the endogenous non-symmetric nature of metal ion charges and redistribution in response to environmental stimuli (Li *et al.*, 2013). (C) Derived from this model, the cationic dummy model mimics valence electrons surrounding the central metal atom by placing cations around it (Peters *et al.*, 2010).

To incorporate metal ions into MD force fields, the MCPB built by Peters *et al* (2010) utilizes the electrostatic plus bonding model (Figure 3.2, Panel B). The metalloprotein force field is fully compatible with the AMBER force fields (AMBER FF) (Peters *et al.*, 2010), the AMBER suite of programmes (Case *et al.*, 2005) and CHARMM force fields (Brooks *et al.*, 1983). The MCPB is comparable to the systematic derivation of force field parameters previously described by Hoops *et al* (1991), modelling both forms of the metalloprotein human cationic anhydrase II (HCAII) via extension of the original AMBER force field to include zinc ions. The Zinc AMBER Force Field (ZAFF) therefore incorporates the MCPB programme parameters (Peters *et al.*, 2010).

In silico exploration of the model's properties in water, including: backbone movement and fold, will provide additional insight into the basic characteristics of the CHFR RING domain. In addition, average fluctuations (or movement) of individual residues within the model, cross-referenced with modelled secondary structural content, will provide insight into behaviour of the model on a residue-by-residue basis.

3.1 MATERIALS AND METHODS

3.1.1 Multiple sequence alignments and molecular visualisation

Multiple sequence alignments serve as predecessors to other computational methods; including secondary structure prediction, identification of genomic arrangements and phylogenetic analysis (Waterhouse *et al.*, 2009). Information concerning both buried and functionally important residues can be extrapolated from conserved regions between variants (Waterhouse *et al.*, 2009; Jones, 1999). For example, Benner and Gerloff (1991) demonstrated (with reasonable accuracy) that solvent accessibility could be predicted through an array of sequences and determining the degree of variability between them.

To facilitate the identification of structurally and/or functionally evolutionary conserved residues within the CHFR-RING domain, two independent multiple sequence alignments were produced. The first included RING domain-containing homologues from a wide range of taxa; providing invaluable information concerning evolutionary conserved and therefore important residues from a wide range of homologs without bias. The second used RING domain protein sequences retrieved exclusively from the protein databank (PDB). In combination with molecular visualisation of crystallographic-derived 3D structures, this provided a layer of experimentally determined secondary structural information, comparable between each of the PDB-derived homologues and the respective CHFR RING domain sequence.

Multiple sequence alignments were performed in JalView (Version 2.8.2; Clamp *et al.*, 2004; Waterhouse *et al.*, 2009) using Clustal Omega (Sievers *et al.*, 2011; Goujon *et al.*, 2010) and manual adjustment; including amino acid sequences obtained from the NCBI nr database or PDB (Abola *et al.*, 1987). To ‘anchor’ the RING domain sequences into a position identical to the CHFR (isoform 1) target, C3HC4-type zinc binding (or ‘cross-brace’) motifs within the protein homologs (C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48)C-X2-C) were first aligned between sequences; serving as a consistent feature between different sequences, whilst maintain the features of other alignments. The percentage identity and hydrophobicity visualization features within JalView further aided the manual editing process of multiple sequence alignments.

Within the PDB-specific alignment, secondary structural information was obtained from PDB files and visualised in Pymol (Version 1.8.0.5; Schrödinger, 2015); with positioning of alpha helices and beta sheets (to respective protein sequences) included via the JalView Annotator tool (Waterhouse *et al.*, 2009).

Pymol (Version 1.8.0.5) was also used for all proceeding protein structure visualisations, protein structural alignments and production of high quality figures (Schrödinger, 2015).

3.1.2 Secondary structure prediction

Usage of multiple sequence alignments in conjunction with secondary structure prediction tools frequently facilitates the observation of deletions, insertions and variability between known and unknown protein structures (Jones, 1999; Aliermann *et al.*, 1978; Zvelebiest *et al.*, 1987). In addition, the evolutionary conservation of residues with similar physiochemical properties serves as a powerful prediction tool regarding protein fold (thus structure) and function (Cuff and Barton, 1999).

Whilst secondary structural information alone is not useful in producing reliable and detailed 3D models (Cole *et al.*, 2008), its prediction is highly important in

homology modelling, usage of fold recognition techniques and *ab initio* 3D structure prediction methods (Drozdetskiy *et al.*, 2015). In addition, secondary structural information is essential in guiding both deleterious mutations and site-directed mutagenesis (SDM) experiments, whilst maintaining overall protein structure (Cuff and Barton, 1999).

The JPred secondary structure prediction server is a three-state prediction tool of alpha helices, beta strands and coils (Drozdetskiy *et al.*, 2015; Cole *et al.*, 2008). Evolutionary information is derived from single protein sequences, a multiple sequence alignment or batches of protein sequences through a Java interface (Drozdetskiy *et al.*, 2015). The integrated JNet 2.3.1 algorithm is a powerful neural network based predictor (Drozdetskiy *et al.*, 2015; Fox *et al.*, 2014). Exploiting the features of 6 individual heuristic algorithms, it enlists a progressive sequence alignment method (via PSI-BLAST; Altschul *et al.*, 1997) and an accurate consensus of secondary structure prediction (Cuff and Barton, 2000; Cuff *et al.*, 1998). JNet standardizes the respective input formats per algorithm, inclusive of conservation values, solvent accessibility, prediction reliability and physiochemical properties (Zvelebil *et al.*, 1987; Cuff *et al.*, 1998; Cuff and Barton, 2000).

PSI-Pred is an alternative secondary structure prediction tool, which also utilises a simplified and accurate two-stage neural network (Buchan *et al.*, 2013); with secondary structure prediction based upon structural, rather than sequence similarity (Jones, 1999).

The three-stage prediction method initially obtains the (intermediate) sequence profiles within the PSI-BLAST log file (Altschul *et al.*, 1997), converting the position-based scoring matrix into a window of 15 rows (Jones, 1999). The scaled inputs are then transferred to the first network; consisting of 75 hidden units, 315 inputs and 3 outputs (Jones, 1999). Following the conversion of these outputs to new windows (15 x 30), they serve as an input to the second (filtering) network; consisting of 60 hidden units, 60 inputs and 3 outputs; proceeded by an

intermediate stage of secondary structure prediction (Jones, 1999). A final stage filters a single three-state secondary prediction report (Jones, 1999).

A highly stringent and blind evaluation of structural prediction methods is employed via the CASP experiments (Anon, 2015); a useful tool to fairly compare computational structure prediction methods (Moult *et al.*, 1997). During the 4th CASP experiment (CASP4, held in 2000), PSI-Pred (Version 2.0) and JPred (Version 2.0) scored an average three-state prediction accuracy (Q_3) of 82.0 and 81.6%, respectively (Drozdetskiy *et al.*, 2015; Anon, 2015). Due to the small sample sizes used, statistically significant results are not obtainable from CASP experiments (Moult *et al.*, 1997). However, the Q_3 values obtained from the experiments do provide some comparative insight into the advancements within different secondary structural tools and their accuracies (Moult *et al.*, 1997).

Taking both the close prediction accuracy and variability of method evaluation (Jones, 1999; Cuff and Barton, 2000) between the methods into consideration; JPred (Version 4; utilizing JNew Version 2.3.1; Drozdetskiy *et al.*, 2015) and PSI-Pred (Version 3.3; Buchan *et al.*, 2013; Jones, 1999) were both independently used to predict the secondary structure of the CHFR RING domain. The JNet Pred prediction from JPred was included as the consensus secondary structure prediction.

Secondary structural predictions were then included within the PDB-specific multiple sequence alignment, with annotations of predicted alpha helices and beta sheets positioned corresponding to the respective CHFR RING domain residues.

3.1.3 Domain boundary prediction

Prior to homology modelling of the CHFR central RING domain, domain boundaries were defined via analysis of predicted secondary structural content predicted (Section 3.1.2); taking care to include full (and non-truncated) alpha helices and/or antiparallel beta strands within models. In addition, the full CHFR sequence was submitted to the InterProScan (Version 5) (Jones *et al.*, 2014, Finn

et al., 2017; Zdobnov and Apweiler, 2001) to scan 5 protein annotation databases, ascertain signature matches with protein families and determine RING domain boundaries and the spanning respective amino acid sequence within the CHFR RING domain.

3.1.4 Homology modelling

3.1.4.1 Basic and multi-template modelling

Homology (or comparative) modelling of protein structures involves prediction of a protein 3D structure based upon an alignment between the target protein sequence and one or more related template (sequence(s)) (Sanchez and Sali, 2000). Prior to deducing the structural basis underpinning the central CHFR RING domain recognition of E2 substrates, a homology (3D structural model) was produced.

In order to obtain 3D models, homology modelling was performed using MODELLER (Version 9.15, Eswarn *et al.*, 2006), an application that facilitates the production of 3D models by the satisfaction of spatial restraints (Sali and Blundell, 1993). Potentially restraining the main chain conformation of modelled residues; the local similarities between sequences within the templates – target alignment, individual residue properties and conformational features of the main chain of corresponding residues within the related protein are collectively expressed as the probability of density function (pdf) and therefore descriptive of features restrained within models produced (Sali and Blundell, 1993).

Templates obtained by X-ray crystallography or NMR which exhibited a high to medium resolution (approximately between 1.0 – 2.5Å, inclusive), medium to high percentage identity to the CHFR RING target (>30%, calculated via submission of the to the SIAS server (SIAS, 2016)), low R_{free} value (<0.250). Where structures of E3 ubiquitin ligases bound to known CHFR E2 substrates were identified (such as

RNF4: UbcH5a: Ub, PDB:4AP4 and RNF8: Ubc13: Mms2, PDB: 4ORH), higher quality datasets were prioritized for homology modelling.

Based on criteria outlined, PDB: 4AYC (corresponding to the structure of RNF8, Mattioli *et al.*, 2012) was selected for basic (single) template modelling; solved by X-ray diffraction and crystallized at 1.90Å (R_{free} : 0.229) with an L-peptide linker.

For basic (single-template) modelling between target and sequence, an alignment between the related template structure and target sequence was performed manually and exported from JalView in a FASTA format; proceed by correct formatted into a Modeller Input alignment file PIR (MODELLER-readable) format (Appendix 2, A2.1) to generate basic models. The alignment file features a line including the protein code (tagged by >P1, followed by the protein code); proceeded by the protein code again (identical to line 1) and 10 colons and spaces containing (all applicable) sequence information; such as first/last residue number and chain code, and the target sequence. Proceeding the target sequence, two full stops were included (..) and an asterisk (*) to indicate to MODELLER the presence of two zinc ions and the end of the sequence, respectively. The section below contains identical information, respective to the template selected, with all necessary information described to enable MODELLER to obtain atomic coordinates from the template structure coordinates within the PDB file.

The python script, containing references to both alignment file (described previously) and template PDB file (containing coordinates), was used in the basic modelling protocol within MODELLER (Appendix 2, A2.2), with 100 models generated. To run the python script and generate models in MODELLER, the following command was entered:

```
python CHFRring_4ayc.py
```

To perform multi-template modelling, the template selection criteria were extended to include 6 PDB files, satisfying the criteria previously described and an alignment file produced (containing all 6 sequences and coordinate references) in an identical manner (Appendix 2, A2.3 to A2.6). To generate a multiple sequence alignment file, the MODELLER **salign()** command was used (Madhusudhan *et al.*, 2006). The **salign()** command reads sequences from all 6 specified PDB files (via the **append_model** command), with initial rough alignments generated by **salign()**, evaluated and improved. The customizable python script was edited to include coordinate (template) PDB file names and specific chains (Appendix 2, A2.3), with the script ran via the following command:

```
python salign.py
```

The output PAP file was inspected briefly for errors (if any) and the target sequence aligned to the template structures using the MODELLER **salign()** command; whilst the **align_block** parameter was set to the number of structures contained within the alignment (6). A pair-wise alignment is requested, ensuring that the multiple sequence alignment does not revert to any existing alignments between individual templates; in addition to a **gap_function** to specifically utilize information derived from the 6 specified sequences. The customizable python script was edited to include coordinate (template) PDB file names and specific chains (Appendix 2, A2.4), with the script ran via the following command line:

```
python align_block.py
```

New models for the CHFR RING target sequence based upon multiple templates were then built based on an alignment performed against the multiple templates (previously described; Appendix 2, A2.5), whereby the `model_multi` file python script was used to generate 100 structures (Appendix 2, A2.6):

```
python model_multi.py
```

3.1.4.2 Model validation and additional loop refinement

Usages of both built-in and external model evaluation resources are exceptionally important in ensuring both programme-specific and independent structural criteria are met (Sippl, 1993; Lüthy et al., 1992).

Following production of models via basic and multi-template modelling methods, restraint violations and errors (if any) were initially assessed within the MODELLER log file produced. The DOPE (Discrete Optimized Protein Energy) method assesses the quality of all atoms within models produced using an optimized statistical potential (Shen and Sali, 2006). The DOPE score is considered the most reliable method in order to obtain native-like models, whereby lower DOPE scores correspond to more native-like models (Shen and Sali, 2006). However, multi-template modelling of the CHFR RING domain revealed a difference in DOPE score of -4 between best (model with lowest DOPE score) and second best (model with second lowest DOPE score) models. The molecular probability density function (molpdf) function is optimized and implemented within the MODELLER programme, whereby optimized models contain a minimum number of violated restraints and are therefore assigned a low molpdf value (Sali and Blundell, 1993). Both molpdf and DOPE scores were also obtained from the log file and therefore taken into consideration when evaluating models produced.

Tabulated molpdf and DOPE scores were then sorted in ascending order using Microsoft Excel for Mac (2011, Version 14.6.2). Whilst models with the lowest DOPE and molpdf values are considered the most accurate, it is not necessarily true that a single model will contain the lowest of both values. Therefore, per model, the molpdf score was added to the (negative) DOPE score and models ranked in ascending order, in addition to cross-referencing relative individual ranks of molpdf and DOPE scores within the best models. Taking this into consideration, models with the lowest molpdf + (-) DOPE score were therefore selected as the most native representation of the CHFR RING domain.

Within the best models, alignment position was plotted as a function of DOPE score, enabling assessment of models on a residue-by-residue basis. An evaluation customizable script was used to produce profiles of DOPE data per model. To run the evaluation python script (Appendix 2, A2.7) in MODELLER, the following command was entered:

```
python evaluate_template4ayc.py
```

In addition, to ensure correct zinc-ligand (residue) coordination was observed within the 'best' models, PyMol was used to visually ensure that zinc ion geometry centres were tetrahedral; with at least 4 ligands (cysteine or histidine residues) coordinated to each zinc ion.

The alignment position-DOPE score plots were then assessed for profile, relative to template used, whereby high-energy regions reported by the DOPE score were targeted for loop refinement, where applicable.

The automated loop refinement technique within modeller significantly improves loop prediction accuracy by an additional step of specified energy optimization, utilizing molecular dynamic, conjugate grade and simulated annealing techniques (Fiser et al., 2000). Predicted loop conformations therefore correspond to the lowest energy conformation attainable (Fiser et al., 2000). The loop refinement protocol includes specifying the regions to refine (within the existing coordinate file) by using the loopmodel MODELLER class (Fiser et al., 2000). To run the loop refinement python script in MODELLER (Appendix 2, A2.8), the following command was entered, specifying the start and end residues to include in the process:

```
python Loop_refine.py
```

Evaluation for loop refined models was then repeated, as previously described, with 500 models produced and significant care taken to identify and evaluate models exhibiting correct coordination of residues with the second zinc ion.

3.1.4.3 External evaluation resources

The ProSA web based server facilitates assessment of protein model quality, identification of potentially problematic areas within protein 3D structures and is used in the validation of models produced by X-ray crystallography, NMR and 3D modelling (Sippl, 1993). Atomic coordinates are supplied via uploading a PDB file or specifying the relevant PDB code, with two important outputs consisting of Z-score and residue energy plots (Wiederstein and Sippl, 2007; Sanchez and Sali, 2000). The Z-score serves as an assessment method of the overall quality of models produced, whereby total energy of protein structures are assessed and deviations measured, with respect to energy distributions attainable from possible random protein conformations (Wiederstein and Sippl, 2007; Sippl, 1993). The Z-score of input models is plotted alongside Z-scores of known NMR and X-ray derived structures; whereby inclusion of the model Z-score within these two regions is indicative of a high quality model, unlikely to contain errors (Wiederstein and Sippl, 2007). ProSA (2003) was therefore used to calculate the Z-score of all models produced (Wiederstein and Sippl, 2007.).

Stereochemical quality of models produced can be assessed at both local and global scales by examining potential ψ and ϕ dihedral angle conformations, whereby distinct clustering of angles correlates with both parameter distribution and higher resolution data sets (Morris *et al.*, 1992). Ramachandran plots therefore indicate energetically allowed regions for both dihedrals within proteins and polypeptides (Ramachandran *et al.*, 1963). Acquirement of such data facilitates identification of unusual stereochemistry and potential sources of experimental error (Morris *et al.*, 1992). ProCheck (Version 3.5, Laskarski *et al.*, 1993) was used to produce Ramachandran plots and identify potential sources of error within protein models produced.

3.4.1.4 Control model

Additional validation of the modelling procedure for 3D prediction of the CHFR RING domain structure was also implemented via a control model. The RNF8 E3 ubiquitin ligase homolog (PDB: 4AYC) was modelled using an identical approach previously described and using the RING domain template of RNF73 (PDB: 2Y43, Huang *et al.*, 2011) in the absence of any known RNF8 structural information. RNF8 deposited structures within the PDB were then compared to the final model via structural alignments (to the known RNF8 crystal structure, PDB: 4ORH) and RMSDs between model and the experimentally determined structure was also calculated (both in PyMol).

3.4.1.5 Iterative modelling via extensions of the CHFR RING: E2 ~ Ub interfaces

In order to predict residues at interfaces within the CHFR RING: Ubc13 (: Mms2) ~Ub and CHFR RING: UbCH5a ~Ub protein complexes, the PDB was searched for crystallographic data sets containing RING E3 (CHFR) homologues in complex with E2s that were verified to interact with the full-length CHFR protein (*in vitro*) (Table 3.1).

Table 3.1 Summary of PDB crystal structures of E3: E2 and E3: E2 ~ ubiquitin complexes used to predict CHFR RING interfaces

	PDB	Citation
RNF8: Ubc13 (: Mms2)	4ORH	Campbell <i>et al</i> (2012)
RNF4: Ubc13 (:Ube2V2) ~ polyubiquitin C	5AIU	Branigan <i>et al</i> (2015)
RNF4 : UbCH5a ~ ubiquitin	4AP4	Plechanovova <i>et al</i> (2012)

The initial CHFR RING homology model cartoon (106) was then aligned (in PyMol) with the respective E3 ubiquitin ligase counterparts within the E3 : E2 ~ubiquitin protein complexes (Table 3.1) using Pymol; with polar contacts identified by extensively searching all interfaces with the CHFR RING homology model. In addition, an annotated multiple sequence alignment (using JalView and Clustal Omega, with manual adjustment; Clamp *et al.*, 2002 and Sievers *et al.*, 2011)

between E2s and the CHFR RING domain informed of secondary structure content within the RING model and respective E3 homologues.

Proceeding identification of interfaces present within the CHFR RING: UbcH5a ~ Ub and CHFR RING: Ubc13 ~ Ub complexes, all respective interactions described within the RNF4 E3 ubiquitin ligase homologue (Plechanovova *et al.*, 2012) were similar or identical (in the majority of cases), with the exception of the Ub G35, RING-1 H160 and RING-2 Y193 stacking interaction. As a result, the original CHFR RING 106 model was extended to include Y362 and L363 (corresponding to Y193 and I194 of RNF4, respectively) within additional basic, multi-template and loop refined models in MODELLER (Eswarn *et al.*, 2006; Sali and Blundell, 1993).

Initial, independent PDB (Abola *et al.*, 1987) BLAST (Altschul *et al.*, 1997) searches using the extended N-terminus of the CHFR RING domain (residues V346 to L363) and full-length extended protein (residues T300 to L363) did not produce any suitable templates. A basic model (using RNF4, PDB: 4AP4 as a template) and a multi-template model were therefore produced. The multi-template model included the original RING CHFR model 106 and RNF4 (PDB: 4AP4, Plechanovova *et al.*, 2012) as templates (with PDB: 4AP4 to 'anchor' the loop and Y193 residue in the correct position during the modelling process). Whilst PDB: 3NG2 (*R. norvegicus*) is a higher quality RNF4 dataset than PDB: 4AP4 (1.8Å and 2.21Å, respectively), the N-terminal loop region is in an incorrect conformation to map the Y-H-G stacking interaction. Therefore, PDB: 4AP4 was chosen as a PDB template.

The modelling procedure was implemented using similar basic and multi-template modelling strategies previously described, with (internal and external) evaluations and loop refinements also enlisted.

3.1.5 Molecular dynamic simulations

Molecular dynamic (MD) simulations involve following the dynamic and structural properties of macromolecules, molecules and proteins in space and time, bridging the experimental gaps between macroscopic kinetic and structural data (Schlink *et al.*, 2010). This could provide valuable information, regarding the dynamic and structural properties of CHFR RING domain model 10. In addition, comparison of the original zinc ion containing homology model with a zinc-deficient counterpart could provide information regarding the behaviour of the zinc metal ion centres and any potential role in stabilizing the solvated RING domain.

MD simulations were performed in GROMACS (Van der Spoel *et al.*, 2005; Pronk *et al.*, 2013). However, prior to implementation of MD simulations, incorporation of metal ions within the GROMACS molecular dynamic force field (FF) parameters (Peters *et al.*, 2010) and specification of ligands (residues) bound to each metal ion centre (ZN1 and ZN2) is required.

The MCPB programme developed by Peters *et al.* (2010) utilises the *electrostatic plus bonded model* (Figure 3.2, Panel B); defining the bond angle between the metal ion and ligand (here, zinc and cysteine/ histidine residues within the CHFR RING domain); specifying the electrostatic potential (ESP) charges present. Partial atomic charge calculations are used to determine the force (angle) constraints, regulated torsion terms and bond angles; incorporating a combination of *ab initio*, experimentally and Density Function Theory (DFT) derived datasets (Peters *et al.*, 2010). The MCPB programme also incorporates the van der Waals (VDW) parameters described by Merz (1991) and Particle mesh Ewald (PME)-simulation and zinc ion specific data produced by Li *et al.* (2013; Peters *et al.*, 2010; Darden *et al.*, 1993). The metalloprotein force field is fully compatible with the AMBER force fields (AMBER FF) (Peters *et al.*, 2010), the AMBER suite of programmes (Case *et al.*, 2005) and CHARMM force fields (Brooks *et al.*, 1983). The Zinc AMBER Force Field (ZAFF) therefore incorporates the MCPB programme parameters and was

used in modelling the CHFR RING metal ion centres, prior to MD simulations (Li *et al.*, 2013; 2015)

Proceeding visualization of metal zinc ion centres and coordinating residues, the model 10 (PDB) file was edited to specify the zinc metal ions centres (ZN1 and ZN2) and their specific coordinated ligand combinations (CCCC and CHCC, respectively) (in accordance with the metal parameterization specified in ZAFF (Peters *et al.*, 2011) (Table 3.2).

Table 3.2 Summary of metal ion centres used to incorporate the ZAFF metal ion parameters of the CHFR RING domain model 10, prior to MD studies (adapted from Li and Merz, 2015).

ZAFF Centre ID	Metal centre type	Zinc ions names		Ligand (coordinated residue) names	
		Current	Edited (ZAFF specific)	Current (residue position)	Edited (ZAFF specific; residue position)
1	Zn-CCCC (four C residues)	ZN	IZN1	C (304)	CY1 (304)
				C (307)	CY1 (307)
				C (328)	CY1 (328)
				C (325)	CY1 (325)
3	Zn-CCCH (three C residues, one H as HID)	ZN	ZN3	C (320)	CY3 (320)
				C (399)	CY3 (399)
				C (342)	CY3 (342)
				H (322)	HD1 (322)

HID: HD1, hydrogen is positioned on the delta electron

As specified by the ZAFF metal ion centre types (Peters *et al.*, 2010, Li and Merz, 2015), histidine residues within PDB files either have a hydrogen on the delta nitrogen (HID), the epsilon nitrogen (HIE) or both (thus a positive charge, HIP). H322 appears to make a polar contact with ZN2 (here, denoted in ZAFF as ZN3, Table 3.2), but since no crystal structure of the CHFR RING domain is available, it is unclear as to whether hydrogen bonds form between this residue and others. Therefore, it is assumed that the hydrogen is position on the delta electron (HID) during preparation of the PDB (coordinate) file of model 10. In addition, termination

(TER) tags were added after each zinc ion to indicate that each ion is independently coordinated to its (residue) ligands.

The AMBER 15 (Assisted Model Building with Energy Restraints) package (Case *et al.*, 2005) was used to produce the topology (.prmtop) and coordinate (.inpcrd) files listing the constant attributes/ outputs and coordinates of each atom in the models, respectively. Specifically, the *tLeap* preparatory program was used to produce both file types, the predecessors for GROMACS-specific topology (.top) and coordinate (.gro) file formats (Case *et al.*, 2005).

Force field parameters of proteins and macromolecules ensure interaction energies and molecular geometries calculated in MD simulations are appropriate, with respect to the specific force field used (Schlink, 2010). Containing information regarding atom types, dihedrals, bonds and angles present, the AMBER Leap 14SB Force Field (ffLeaprc.ff14SB, Maier *et al.*, 2015) and TIP3P water model (Jorgensen and Tiradorives, 1988) was used within *tLeap* to build models, including ZAFF metal ion centre information within the script used (Appendix 2, A2.9). Topology and coordinate files (.prmtop and .inpcrd, respectively) were then converted to the correct GROMACS file formats (.top and .gro, respectively) using AnteChamber Python Parser InterfacE (ACPYPE) within the Linux terminal (and the ACPYPE directory) (da Silva and Vranken, 2012):

```
acpype -p RING_10_GMX.prmtop -x RING_10_GMX.inpcrd
```

The .top file contains force field parameters and information regarding the CHFR RING topology, whilst the .gro file details descriptions of atomic coordinates and (eventual) periodic box dimensions; both now in the correct format for usage in GROMACS.

To obtain a more correct atomic arrangement and improved atomic relaxation, MD simulations were then implemented within the GROMACS package (Version 4.6.5).

The initial system was described within a triclinic unit cell (Abraham *et al.*, 2015), encompassing all periodic constraints (der Spoel *et al.*, 2005). Prior to solvation, optimal box dimensions were explored; to ensure that the unit cell contained minimal water molecules used to solvate the molecule. A range of different box shapes were trialled, with the protein centred and at least 1.0nm from the box edge (therefore at least 2.0nm distance between the two protein periodic images):

```
gmx editconf -f RING_Zn10_GMX.gro -bt dodecahedron -d 1.0 -o
    / box.gro
gmx editconf -f RING_Zn10_GMX.gro -bt cubic -d 1.0 -o box.gro
gmx editconf -f RING_Zn10_GMX.gro -bt octahedron -d 1.0 -o
    / box.gro
```

The dodecahedron, cubic and octahedron boxes contained volumes of 283.70, 2669.60 and 307.51nm³, respectively; therefore the dodecahedron triclinic unit cell was deemed most appropriate (and efficient) for MD simulations and was solvated with water:

```
gmx solvate -cp box.gro -cs spc216 -o RING_Zn106_GMX.gro -p
    / RING_Zn106_GMX.top
```

Differences between the Simple Point Charge (SPC) (Berendsen *et al.*, 1981) and Transferable Intermolecular Potential 3P (TIP3P) (Jorgensen *et al.*, 1983) three-point water models were explored by Mark and Nilsson (2001). Under identical experimental conditions, a faster diffusion was observed via the calculated self-diffusion coefficient of SPC with reduced structure (Mark and Nilsson, 2001). In contrast, even faster diffusion dynamics and loss of additional structural parameters were observable in the TIP3P model (Mark and Nilsson, 2001).

Taking this into consideration, spc216.gro was used as the solvent configuration for the TIP3P water model (Jorgensen and Tiradorives, 1988), since it is applicable

to all 3-point water models (Apol *et al.*, 2010); with 2973 TIP3P water molecules added to the system.

Prior to neutralising the system with ions, the TIP3P water model (from the AMBER 03 Force Field, amber03.ff) and metal ion information (including name, atoms, angles and bond lengths) was specified within the topology file (to ensure atom names and parameters from AMBER were transferred). Within the topology file, conditional statements were included before specification of the [system] content (protein and solvent):

```
; Include water topology
#include "amber03.ff/tip3p.itp"
; Include topology for ions
#include "amber03.ff/ions.itp"
```

And also just prior to specifying the [atom types] within the system:

```
#include "amber03.ff/ffnonbonded.itp"
#ifdef POSRES
#include "amber03.ff/ffnonbonded.itp"
#endif
```

The net charge of the system (as specified in the last line of the [atoms] directory within the topology file) was -6.00, thus the genion GROMACS tool (Apol *et al.*, 2010) was used to replace 6 TIP3P water molecules with 6 sodium (NA) ions within the system and to neutralize it. This required the production of a run input file (ions.tpr) by the GROMACS (GROMacs pre-processor); processing both the RING_10_GMX.gro (coordinate) file and respective topology counterpart to produce an atomic-level input (.tpr) file; containing parameters required for all system atomic components (Apol *et al.*, 2010). The Molecular Dynamics Parameter File for ions (Appendix 2, A2.10) was used to assemble the correct parameters

specified (utilizing both topology molecule information and coordinate information) and assemble the .tpr file:

```
gmx grompp -f ions.mdp -c RING_Zn10_GMX.gro -p  
  \ RING_Zn106_GMX.top -o ions.tpr
```

The output ions.tpr file (containing a system atomic-level description, Apol *et al.*, 2010) was then served as the input file for genion:

```
gmx genion -s ions.tpr -o RING_Zn10_GMX.gro -p  
  \ RING_Zn10_GMX.top -neutral
```

At the prompt, group 19 (“SOL”) was selected for embedding ions, since this instructs genion to replace the solvent (water) molecules and not protein with the appropriate number of ions (Apol *et al.*, 2010). The `-neutral` option specifies to genion to neutralise the system; with the `[molecules]` directive now containing 6 NA ions.

The electroneutral, solvated system assembled was then equilibrated, prior to beginning the production (or dynamic) phase of the simulation. To ensure a stable system is produced, the total, kinetic and potential energy terms are converged during the energy exchanges between kinetic and potential energies (Schlink, 2010). The process also ensures that no incorrect geometry or steric clashes are present, in addition to relaxation of the system (Apol *et al.*, 2010). The Energy Minimization (EM) process began with assembly of the binary input using grompp and the energy minimization input file (Appendix 2, A2.11):

```
gmx grompp -f minim.mdp -c RING_Zn10_GMX.gro -p  
  \ RING_Zn10_GMX.top -o em.tpr  
gmx mdrun -v -deffnm em
```

To summarize, the system was subjected to energy minimization using the steepest descent algorithm (Apol *et al.*, 2010), with a maximum of 50,000 steps used to reach a maximum force (F_{\max}) of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$, which was deemed the most appropriate gradient to produce relaxed models. The EM was then evaluated (for success) by assessing for a negative potential energy (E_{pot}) to the order of $10^5 - 10^6$ (corresponding to the F_{\max} of $<1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$); indicating that the system was stable enough to proceed with simulations. The EM steps were then plotted as a function of energy (with terms contained within the output em.edr file) using xmgrace (Version 5.1.23):

```
gmx energy -f em.edr -o potential.xvg
```

To equilibrate the solvent and ions around the protein, the initial phase was conducted under a canonical (or isothermal-isochoric), NVT (constant Number of particles, Volume and Temperature, Apol *et al.*, 2010) simulation for 500ps using grompp, mdrun and an nvt.mdp ensemble parameter (input) file (Appendix 2, A2.12):

```
gmx grompp -f nvt.mdp -c em.gro -p RING_Zn10_GMX.top -o  
          \ nvt.tpr  
gmx mdrun -deffnm nvt
```

To summarise, the system was heated from an initial temperature of 0 to 300K (within the constant volume) for 500ps and in 50,000 steps; with coordinates, velocities and energies saved every 0.2 ps. Temperature progression was then analysed, with time (ps) plotted as a function of temperature (K, with a 10-point running average calculated) in xmgrace:

```
gmx energy -f nvt.edr -o temperature.xvg
```

The stability of the system was assessed by maintenance of temperature at the target value (300K), prior to progression to pressure and density equilibration. The equilibrated and temperature-stabilized system was then equilibrated for both parameters under the NPT (constant Number of Particles, Pressure and Temperature, or ‘isothermal – isobaric’; Apol *et al.*, 2010) ensemble to closely resemble experimental conditions (Appendix 2, A2.13):

```
gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p
           \ RING_Zn10_GMX.top -o npt.tpr
gmx energy -f npt.edr -o pressure.xvg
gmx energy -f npt.edr -o density.xvg
```

Both pressure and density equilibrations were carried out simultaneously at 300K, in 250,000 steps; with Parrinello-Rahman (NPT) isotropic pressure coupling (Apol *et al.*, 2010). Exponential relaxation of temperature and pressure coupling were at 0.1 and 2.0ps, respectively. Short range van der Waals and electrostatic cut-off values (1.0nm for both) were also used. Coordinates, velocities and energies were saved every 0.2 ps. Time (ps) was plotted as a function of pressure or density using xmgrace, with a 10-point running average calculated. Whilst some pressure fluctuations within the MD simulation are expected, the average density over the course of equilibration was compared to the TIP3P water model experimental value of $1002.00 \pm 1.00 \text{ kg m}^{-3}$ (Mahoney and Jorgensen, 2000).

On completion of system equilibration, production runs were carried out (using an MD-specific script; Appendix 2, A2.14) in 35,000,000 steps (70 ns), with constant pressure, isotropic scaling and 2.0ps relaxation time. 70 ns was deemed a long enough simulation to allow sufficient time for the MD trajectory to potentially stabilize; whilst efficient enough to run to completion within the ARC HPC 48 hour simulation time restriction. Coordinates, velocities and energies were saved every 10.0ps for 5000 steps. Short range van der Waals and electrostatic cut-off values

(1.0nm for both) were identical to those previously used during equilibrium, and the temperature maintained at 300K:

```
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o
      \ md_0_1.tpr gmx mdrun
-deffnm md_0_1
```

Trajectory conversion was performed in GROMACS *trjconv* to correct for periodicity, such as protein diffusion across the unit cell (Apol *et al.*, 2010), selecting the system output when prompted:

```
gmx trjconv -s md_0_1.tpr -f md_0_1.xtc -o md_0_1_noPBC.xtc -
      \ pbc mol -ur compact
```

The root mean squared deviation (RMSD) was then calculated relative to the CHFR RING model (10) protein backbone present in the minimized, equilibrated system; calculated using both the reference structure and RMSD matrices (Apol *et al.*, 2010). The GROMACS built-in RMSD utility (*rms*) was then used to perform additional calculations, selecting the protein backbone for output (when prompted):

```
gmx rms -s md_0_1.tpr -f md_0_1_noPBC.xtc -o rmsd.xvg -tu ns
```

The compactness of a protein structure is approximate to the radius of gyration (R_g , Apol *et al.*, 2010). To calculate this, the following command was issued:

```
gmx gyrate -s md_0_1.tpr -f md_0_1_noPBC.xtc -o gyrate.xvg
```

Protein stability (and therefore maintenance of fold) was evaluated, in terms of steady value of R_g with progression of time (and *vice versa*).

RMS fluctuations (RMSF) are indicative of the standard deviations present of atomic positions within the trajectory, fitted to a reference frame. Average RMSF values were calculated for each (whole) residue within the RING CHFR model 10:

```
gmx rmsf -f md_0_1.xtc -s em.tpr -o rmsf-per-residue.xvg -ox
    \ average.pdb -oq bfactors-residue.pdb -res -aniso
```

xmgrace and MS Excel were then used to plot residue position as a factor of RMSF, with areas of greatest fluctuation (and therefore greater local backbone deviations) identified by calculating and comparing the standard deviation in RMSF values for loops and secondary structures.

Finally, PDB (structure) files were produced of the system:

```
gmx trjconv -s RING_Zn10_GMX.gro -f RING_Zn10_GMX.gro -o
    /solvated_compact_ionsEM.pdb -pbc atom -ur compact
```

3.1.6 Computational resources

All computational work was performed using Advance Research Computing Nodes 1 and 2 (ARC1 and ARC2), part of the High Performance Computing (HPC) facilities at the University of Leeds (UK).

Operating a Linux-based HPC service (using a CentOS 5 operating system), hardware includes a Sin Microsystems x84-64 based server and storage. Per MD simulation run, 16 nodes were used (of 32 available at any given time); with each node containing a single 32GB, with a quad-core Intel X5560 (2.8GHz) processor.

3.2 RESULTS

3.2.1 Comparison of RING isoform domain architecture

According to genome annotations (Ensembl, Release 84; Flicek *et al.*, 2014), the CHFR protein (UniProt: Q96EP1) is encoded by one gene located on the reverse strand of chromosome 12 (132,822,187-132,956,304). The amino acid sequence of isoforms was retrieved from the National Centre for Biotechnology Information

(NCBI) non-redundant (nr) protein sequence database and UniProtKB (Apweiler *et al.*, 2010, Bairoch *et al.*, 2007).

All CHFR RING domains are identical in sequence, with full-length isoform 2 the second largest of 5 produced via alternative splicing. Its sequence considered 'canonical' within the UniProtKB database (Bairoch *et al.*, 2007). It is therefore used in reference to functional residue numbering position information and proceeding CHFR RING domain structural prediction experiments.

3.2.2 Multiple sequence alignment of CHFR homologs

In order to observe sequence conservation between different RING domain containing homologues, an initial BLASTP search (E-value cut off 10.0, BLOSUM62 scoring matrix with masked low complexity) was performed to retrieve homologous sequences. Local and global alignments performed with BLOSUM62 scoring matrices exhibit a better performance; particularly in secondary structure prediction experiments (Elofsson *et al.*, 2002). The FASTA sequence included residues spanning the approximate location of the central RING domain (300-346) of CHFR isoform 1; in accordance to the NCBI database annotations (Accession: NP_001154816) and encompassing the entire RING-finger motif (CxxCx(6)Cx(5)CxxCx(10)CxxCx).

From 20,000 output sequences, the majority consisted of higher primates or vertebrates, with ~90% of proteins hypothetical or predicted. Selection of highly homologous sequences, incorrect assumption of highly conserved residues and extrapolated functional and/or structural significance would result in significant bias within the alignment. To avoid this, sequences from a wide range of taxa were included within the initial manual multiple sequence alignment, with residues from homologues aligned to the human CHFR target according to both percentage identity (Figure 3.3) and hydrophobicity (data not included).

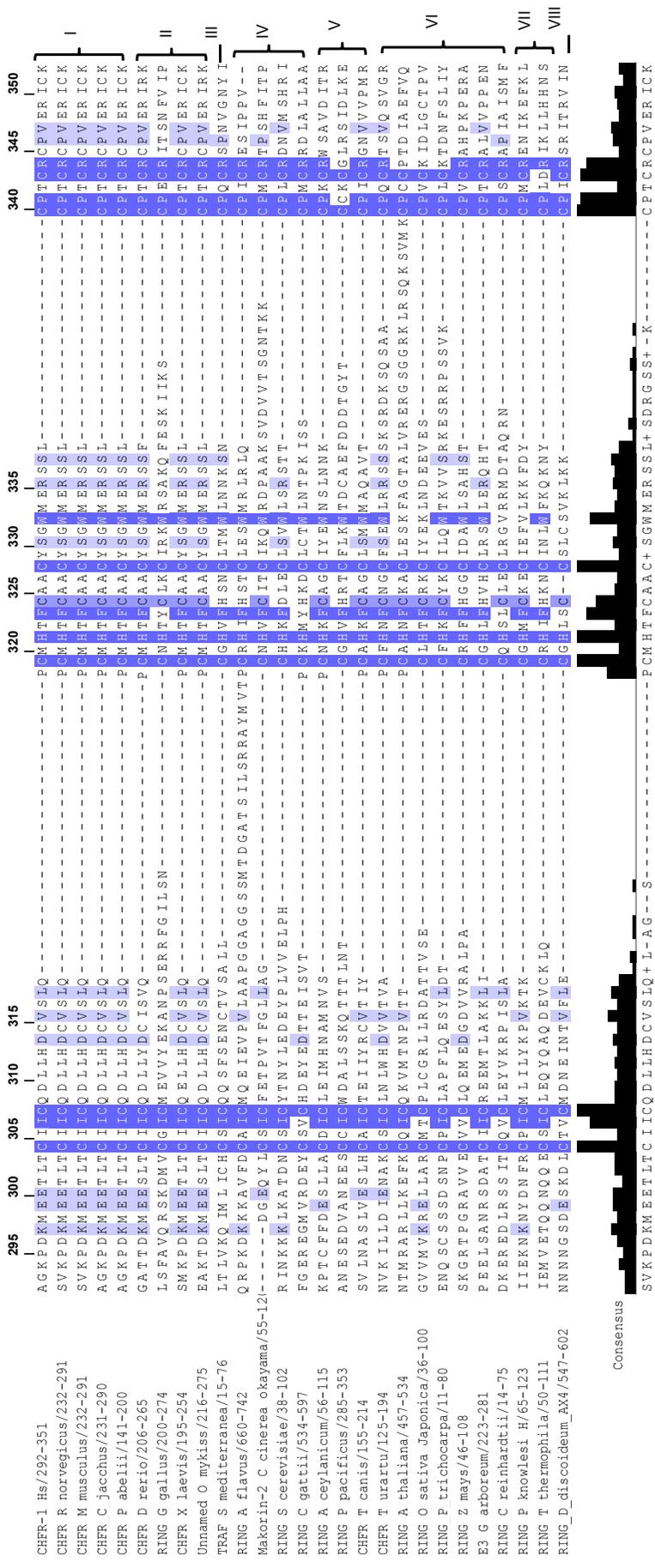


Figure 3.3 Multiple sequence alignment between human CHFR (isoform 1) and other closely related CHFR homologues and RING domain containing proteins. Whilst all mammals exhibit high sequence identity preceding the N and C-terminus of the zinc-binding motif (K297 to T303 and R343 to 351, respectively; human CHFR); such sequence conservation is absent from the majority of other taxa. Sequence colouration corresponds to identical (dark purple) to partially conserved (light purple) residues between sequences. Residue numbers correspond to full-length human CHFR isoform 1.

Concerning criteria for inclusion within the MSA, model organisms were prioritized; since phylogenetic, genomic and proteomic data is most likely to be abundant from these sources. In addition, the most up-to-date NCBI entries were included (where applicable); with theoretical and putative CHFR and RING domain containing homologues excluded. Wide ranges of protein sequences were incorporated, including non-CHFR E3 ubiquitin ligases and RING domain containing proteins.

The consensus zinc binding motif (or RING finger brace motif, Freemont *et al.*, 1991) of CxxCx(10-37)CxHxx(C/H)xxCx(10-29)CxxC was identified in the majority of CHFR and RING-domain containing proteins of organisms from the wide range of classes, clades, subkingdoms and phylum detailed (except *S. cerevisiae*) (Figure 3.3).

3.2.3 Template selection for homology modelling and RING domain boundary prediction

Modelling by satisfaction of spatial restraints (Sali and Blundell, 1993) requires the provision of coordinates derived from a template structure, whereby high-resolution (Hubbard and Blundell, 1987) template(s) with at least 40% sequence identity to the target (Venclovas *et al.*, 1997) are desirable features in obtaining accurate models containing minimal errors. In order to identify a suitable template for homology modelling, a BLAST search was implemented using the PDB (Abola *et al.*, 1987) BLAST (Altschul *et al.*, 1997) tool (E-value cutoff: 10.0, low complexity masked), using residues (C304-C344, human CHFR) encompassing the entire RING domain consensus previously identified. RING domain homologues were sorted according to descending percentage identity (to human CHFR); with information regarding resolution, R-free and R-work values tabulated (Table 3.3).

Table 3.3 CHFR RING domain homologous proteins within the local (PDB derived) database

Name	PDB	Chain	Residue range	Identity (%)	Resolution (X-ray diffraction)	R-free (F)	R-work (W)	F-W
E3 ubiquitin protein ligase: RNF8 (<i>Homo sapiens</i>)	4AYC	A	52-97	41.30	1.90	0.229	0.200	0.029
RING finger protein: RNF165 (<i>H. sapiens</i>)	5D0I	A	41-89	36.17	1.90	0.200	0.173	0.027
E3 ubiquitin protein ligase: RNF2 (<i>H. sapiens</i>)	4S3O	B	49-96	36.17	2.00	0.237	0.194	0.043
E3 ubiquitin protein ligase: RNF146 (<i>Mus musculus</i>)	4QPL	A	4-49	34.78	1.90	0.222	0.186	0.036
RING finger protein: RNF4 (<i>Rattus norvegicus</i>)	3NG2	A	9-61	34.04	1.80	0.242	0.187	0.055
RING finger protein: RNF56 (<i>H. sapiens</i>)	3ZNI	A	336-382	34.04	2.21	0.211	0.175	0.036
RING finger protein: RNF4 (<i>H. sapiens</i>)	4AP4	A	6-58	34.04	2.21	0.242	0.207	0.035
RING finger protein: RNF4 (<i>H. sapiens</i>)	4AP4	A	71-123	34.04	2.21	0.242	0.207	0.035
E3 ubiquitin protein ligase: RNF55 (<i>H. sapiens</i>)	1FBV	A	331-377	34.04	2.90	0.262	0.227	0.035
E3 ubiquitin protein ligase: RNF106 (<i>H. sapiens</i>)	3FL2	A	51-97	31.91	1.75	0.250	0.183	0.067
E3 ubiquitin protein ligase: RNF38 (<i>H. sapiens</i>)	4V3K	C	23-71	31.91	2.04	0.223	0.181	0.042

Table 3.3 CHFR RING domain homologous proteins within the local (PDB derived) database

Name	PDB	Chain	Residue range	Identity (%)	Resolution (X-ray diffraction)	R-free (F)	R-work (W)	F-W
E3 ubiquitin protein ligase: RNF73 (<i>H. sapiens</i>)	2Y43	A	21-67	27.65	1.80	0.223	0.175	0.048
E3 ubiquitin protein ligase: Parkin (<i>H. sapiens</i>)	4I1F	A	274-313	25.00	1.58	0.245	0.205	0.040
E3 ubiquitin protein ligase: RNF31 (<i>H. sapiens</i>)	5EDV	A	3-59	23.40	3.48	0.303	0.249	0.054
E3 ubiquitin protein ligase: RNF31 (<i>H. sapiens</i>)	4LJO	A	18-56	23.07	1.56	0.212	0.181	0.031
E3 ubiquitin protein ligase: RNF106 (<i>H. sapiens</i>)	3T6R	A	6-48	16.27	1.95	0.229	0.194	0.035
E3 ubiquitin protein ligase: RNF106 (<i>H. sapiens</i>)	3ZVY	A	3-45	16.27	1.95	0.227	0.181	0.046
E3 ubiquitin protein ligase: RNF106 (<i>H. sapiens</i>)	3SHB	A	9-51	16.27	1.80	0.228	0.210	0.018

Sequence identity to the human CHFR RING domain was calculated via submission of the MSAs (Figure 3.3, within a FASTA format) to the Sequence Identity And Similarity (SIAS) server (BLOSUM62 scoring matrix; SIAS, 2016).

The majority of protein structures within the local database were solved at a medium to high resolution, with percentage sequence identity to the target ranging from 16.27 to 41.3% (Table 3.3). As a representative sample of RING domain containing proteins and E3 ubiquitin ligases, the local database is therefore suitable for predicting secondary structural features and corresponding predictable functions CHFR RING domain.

Human RNF8 (PDB: 4AYC) was selected as the best template for the single-template stage of homology modelling, considering the low E-value ($1.60093E-4$, data not shown) obtained from the BlastP search; indicative of a statistically significant match of the sequence to the template during the database search. In addition, the template has the highest percentage identity of all homologues to the human CHFR target sequence (41.30%), and high resolution of the crystallographic dataset (1.9Å), relative to all other PDB structures. For multi-template homology modelling, the top 6 PDB structure datasets (sorted from highest to lowest sequence percentage) were selected, with percentage identities to human CHFR ranging from 34.04 - 41.30% and medium to high resolutions evident (2.21 – 1.80Å).

To further facilitate prediction of the CHFR RING domain secondary structural elements and further evaluate its content within the proceeding models, JPred (Drozdetskiy *et al.*, 2015) and PSI-Pred (Buchan *et al.*, 2013) were both independently used to predict the secondary structure of the CHFR RING domain. Annotations derived from these tools were included within a second multiple sequence alignment produced; manually aligning the selected 6 CHFR homologues within the local database with respect to the human CHFR target according to residue percentage identity (Figure 3.4) and hydrophobicity (data not

shown). In order to extrapolate additional predictive information, regarding CHFR RING secondary structure content, secondary structural elements of all PDB files (as identified via visualization of structures in Pymol) within the local database were also included as annotations within the multiple sequence alignment. Within the manual alignment, care was taken to avoid gaps within buried regions (informed by hydrophobicity annotation tools) and secondary structure elements (Sanchez and Sali, 2000).

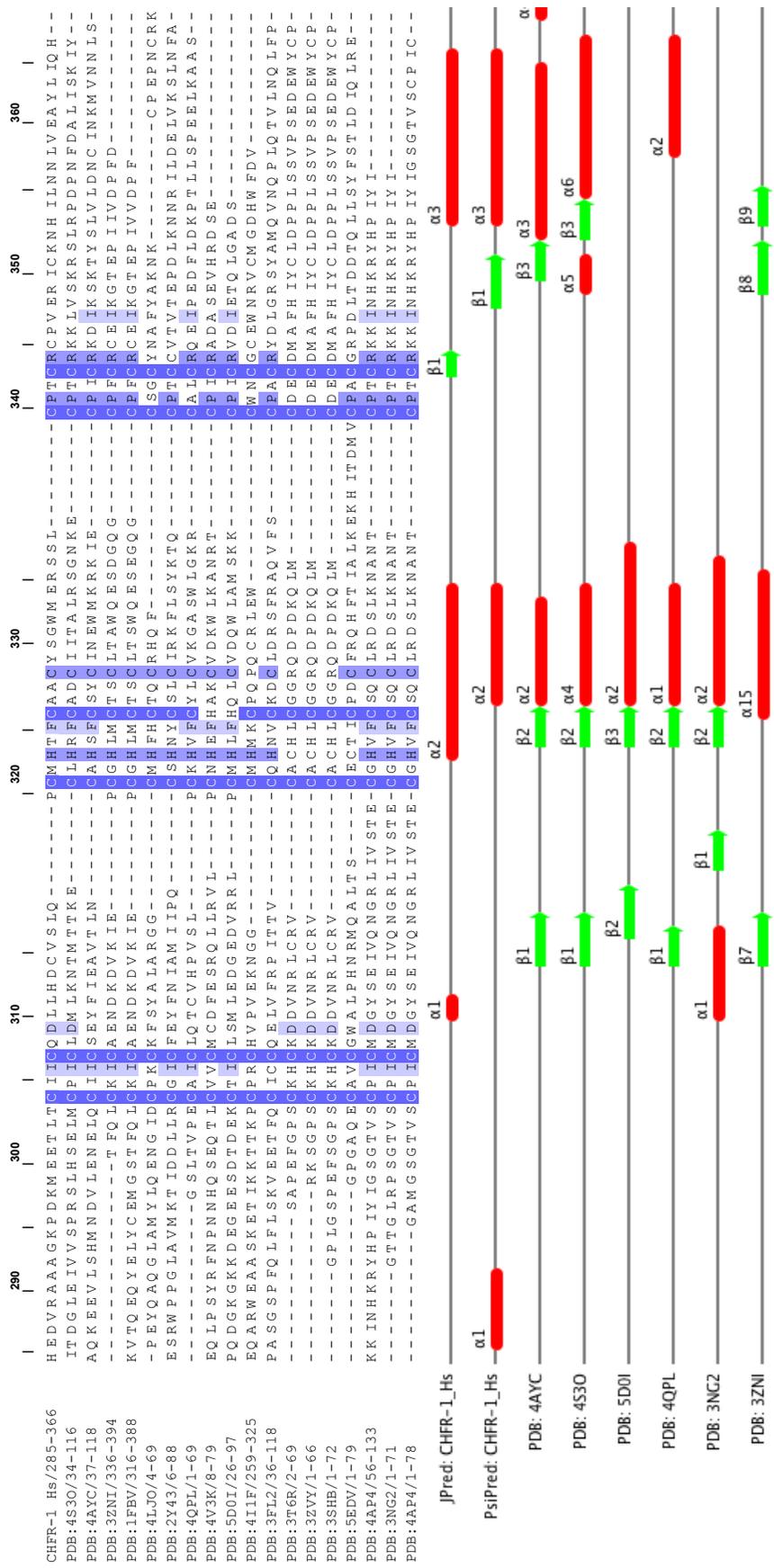


Figure 3.4 Multiple sequence alignment between human CHFR (isoform 1) and the local database structures. All homologues within the local PDB database are included within the main alignment. Secondary structural annotations regarding Jpred and PsiPred predictions, in addition to selected PDB structural datasets, are included as alpha helices (red lines) and beta strands (green arrows). Sequence colouration corresponds to identical (dark purple) to partially conserved (light purple) residues between sequences whilst residue numbers correspond to full-length human CHFR Isoform 1.

Structural homologues and the RING domain target contain the conserved ZINC binding motif or slight variation, previously described (here, CxxCx(11-16)CxHxx(C/H)xxCx(4-18)CxxC).

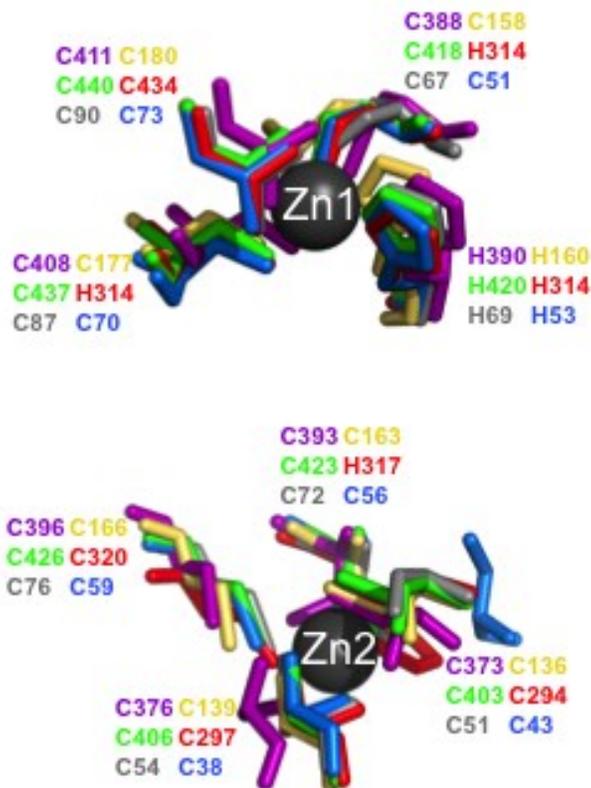
In addition, a strong consensus for a predicted central helix within the CHFR RING domain (at least spanning residues A326 to M333) is shared between JPred and PsiPred secondary structure prediction tools and all 6 PDB structures. Considering alanine (A), glutamic acid (E), leucine (L) and Methionine (M) are preferential to alpha helix formation, this corresponds to approximately 44% of residue content. Within this region, side chains also change from hydrophobic to hydrophilic (with a periodicity of 3 residues), possibly corresponding to the predicted 3.6 residues per turn within the helix.

Within the zinc binding domains, the majority of PDB structures (4AYC, 4S3O, 5D0I 4QPL and 3NG2) exhibit a beta strand prior to the central alpha helix; corresponding to a consensus sequence of HxF. Considering the CHFR RING domain also contains this sequence (H322 – F324), there is a strong possibility that the beta strand is also present in the target. Interestingly, all 6 selected PDB structures contain a hairpin loop connecting the first beta strand to a additional one; corresponding approximately to the positions of D313 to Q318 of the CHFR RING target. The poor sequence identity between PDB structures and target, coupled with a lack of consensus concerning residue type and strand position, suggest the presence of an additional beta strand within the CHFR RING domain is uncertain.

Residues located at the (target) CHFR RING domain N-terminus (H285 to T304) and C-terminus (C344 to Q365) exhibit extremely poor sequence identity to all local database PDB files, with significant variability in secondary structure content (or lack thereof) evident, strongly suggesting loops encompass these regions.

Taken together, the multiple sequence alignments, secondary structure data from PDB files and predicted secondary structures (from JPred and PsiPred) indicate a beta strand (H322 – F324) and a central alpha helix (H322 – F324) are highly likely to be present within the CHFR RING domain. In addition, an N-terminal beta strand (anywhere between L310 and Q318) may also be present within the RING domain.

Prior to production of additional multiple sequence alignments necessary for the proceeding homology modelling, prediction of the CHFR RING domain boundaries was essential. Taking into consideration both the sequence of the zinc-binding motif (containing a total of 7 cysteine residues and 1 histidine residue), the metal-ligand (residue) centre geometry is most likely to be tetrahedral and bound to at least two zinc ions. In addition, this is also the consensus zinc ion centre geometry between the 6 high-identity template structures to be used for multi- template homology modelling (Figure 3.5).



PDB: 3ZNI **PDB: 3NG2** **PDB: 4S3O**
PDB: 4AYC **PDB: 5D0I** **PDB: 4QPL**

Figure 3.5 Zinc ion geometry centres of protein homologues used for basic (single template, PDB: 4AYC) and multi-template (PDBs: 3ZNI, 4AYC, 5D0I, 3NG2, 4S3O and 4QPL) homology modelling of the CHFR RING domain. Coordinating cysteine and histidine residues are shown as sticks, with zinc ions shown as dark grey spheres.

Prior to defining the exact boundaries corresponding to the CHFR RING domain, the amino acid sequence of the full-length CHFR protein was submitted to InterPro (Jones *et al.*, 2014; Finn *et al.*, 2017); utilizing multiple annotation databases and to inform classification according to specific protein domains attributed to distinct families and functional sites (Table 3.4).

Table 3.4 Prediction of CHFR RING domain amino acid boundaries using InterProScan (Zdobnov and Apweiler, 2010). InterPro identifiers, per domain, are included in brackets.

	FHA	Zinc-finger/ RING type	Zinc finger, RING/FYVE/PHD-type
Domains and repeats	13-117 (IPR000253)	284-398 (IPR001841)	489-659 (IPR013083)
Detailed signature matches	13-117 (SSF49879) (SMAD/FHA domain; superfamily)	304-342 (SM00184) 304-343 (PS50089)	

Taking into consideration both Domains and Repeats and Detailed Signature Match annotation hierarchies, the (arithmetic mean) average domain boundaries of the CHFR central RING domain using InterProScan would include residues 297 to 361.

Long loop regions utilize such metal ions via maintaining coordination (of 4 residues per zinc ion), providing protection from proteolytic cleavage and additional stability. In addition, sequence conservation between templates was taken into consideration to maximize the quality of the final homology model. It was therefore predicted that at least residues E300 to V346 would provide sufficient additional loop regions necessary to stabilize the central zinc binding domain of RING models produced and include the majority of residues within InterProScan consensus.

3.2.4 Homology modelling

3.2.4.1 CHFR RING domain

Initially, the Basic Modelling (single template) strategy was used to obtain a 3D model of the CHFR RING domain using MODELLER (Eswarn *et al.*, 2006; Sali and Blundell, 1993). Alignment between the target (CHFR RING domain) and the template structure of RNF8 (PDB: 4AYC) were used to produce 100 models, with best model identified via evaluation of MODELLER-derived DOPE and molpdf

scores, representative of the calculated optimized statistical potential (Shen and Sali, 2006) and the sum of all restraints (Sali and Blundell, 1993), respectively. The strategy was repeated for multi-template and loop-refinement modelling methodologies; with the lowest DOPE and molpdf scores tabulated and DOPE/alignment position data plotted (Table 3.5, Figure 3.6) and corresponding to the respective best energy models.

Table 3.5 Evaluation of CHFR RING domain homology models produced in MODELLER.

Strategy	Best model (/of)	molpdf	DOPE	molpdf+(-)DOPE
Basic (single template)	91 (/100)	1852.666	-4294.051	-2441.385
Multi-template	39 (/100)	1782.120	-4004.930	-2222.811
Loop refinement	106 (/500)	21.156	-4378.275	-4357.119

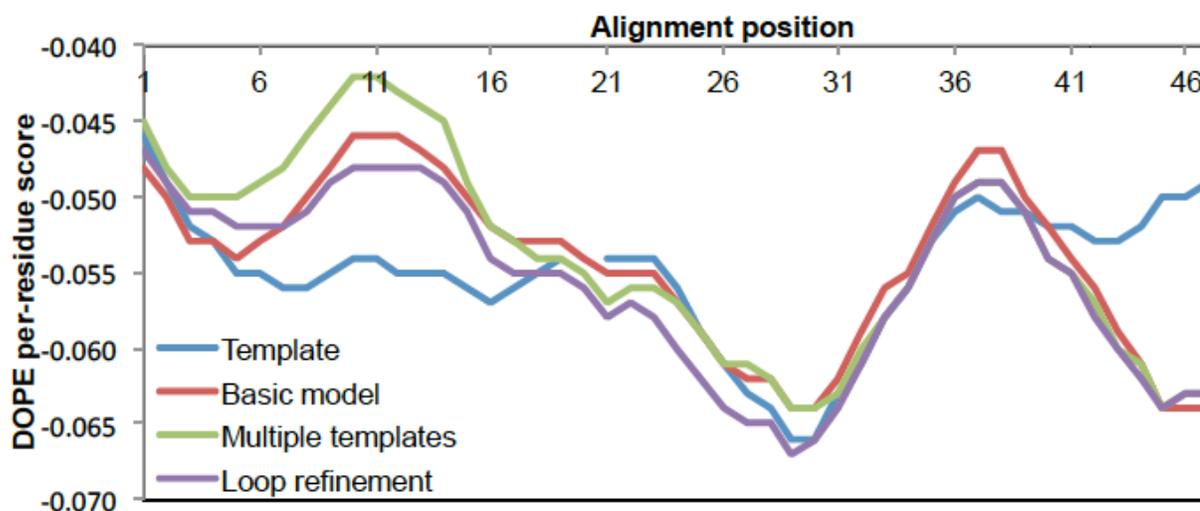


Figure 3.6 DOPE score profile of the CHFR RING model and template (PDB: 4AYC) obtained in basic, multi-template and loop refinement modelling strategies in MODELLER. The DOPE profile of the 4AYC template is shown in blue, with respective colours representative of the proceeding basic, multi template and loop refinement stages. The `evaluate_template4ayc.py` script (Appendix 2, Figure A2.7) was adapted (per model) accordingly.

As identified by the DOPE profile of multi-template model 39 (Table 3.5 and Figure 3.6), residues spanning the CHFR RING (target) and PDB: 4AYC (template) alignment position 6 to 16 (inclusive) exhibited a higher energy profile than the

basic model counterpart, indicating requirement of *ab-initio* loop refinement within this region of the model. Multi-template model 39 was therefore selected as the best model for additional loop refinement.

From 500 models produced by the loop refinement strategy, ~10% exhibited correct zinc ion (2) and ligand (CHCC) coordination, and were therefore selected for additional evaluation; with DOPE and molpdf scores tabulated and DOPE/alignment position data plotted (Figure 3.7) and corresponding to the best energy models.

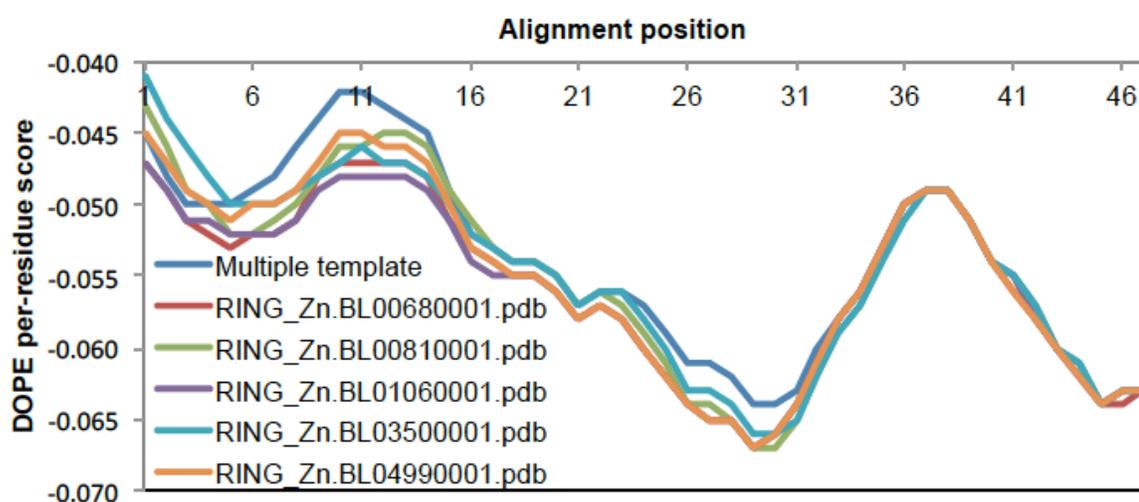


Figure 3.7 DOPE score profile of the CHFR RING models loop-refined between alignment positions 6 to 16 in MODELLER. The `Loop_refine.py` and `evaluate_template4ayc.py` scripts (Appendix 2, A2.8 and A2.7, respectively) were adapted accordingly.

Fluctuations in loop conformations between models, spanning alignment position 6 to 16, can be viewed in the cartoon of the top-5 aligned loop refined models (Figure 3.8).

CHFR-RING (Model: 106)

Other models

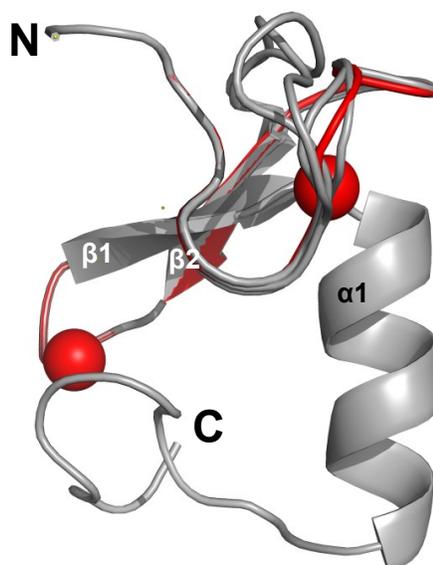


Figure 3.8 Superimposition of the top 5 calculated loop conformations of the CHFR RING domain, refined in MODELLER using the multi-template derived model 39. The best loop-refined model (exhibiting the lowest DOPE profile, spanning alignment position 6 to 16, Figure 9) was identified as Model 106 (Shown in red). Other models with higher molpdf and DOPE scores are shown in grey.

With respect to the original multi-template model (39), loop refined model number 106 exhibited the lowest DOPE profile and was therefore considered the best CHFR RING model. To observe secondary structure content and correct zinc ion coordination between zincs (1 and 2) and respective (cysteine and histidine) ligands, CHFR RING model 106 was viewed as a cartoon in PyMol and aligned to the original PDB: 4AYC template. In addition, a multiple sequence alignment with secondary structure annotations further informed the secondary structure comparison between 4AYC (RNF8) template and CHFR RING target (Figure 3.9).

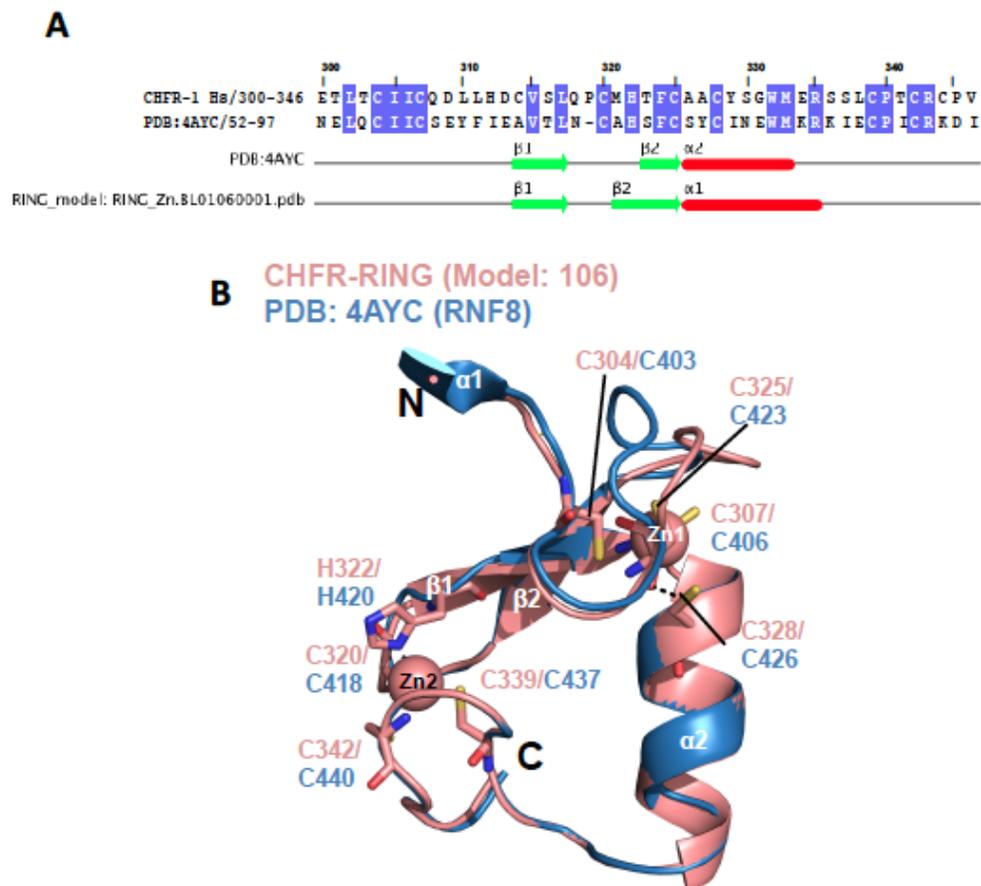


Figure 3.9 Multiple sequence and cartoon alignment between template (PDB: 4AYC, RNF8) final model (106) of the CHFR RING domain, produced by homology modelling using MODELLER. (A) A multiple sequence alignment between PDB template 4AYC and CHFR RING model 106 indicates a high sequence identity (dark purple) within the majority of residues. Within model 106, a central alpha helix (red line) spans residues A326 to R335, preceded by two antiparallel beta strands (C314 to L317 and M321 to C325, respectively; green arrows); both present in the PDB: 4AYC template. (B) Alignment of both model and template indicate identical positioning of both antiparallel β strands (1 and 2), with the second CHFR β strand slightly longer than the template equivalent. The N-terminal α -helix of the 4AYC PDB template is absent from the CHFR RING model. Alpha helical numbering corresponds to the 4AYC PDB template.

As predicted by both JPred and PsiPred secondary structure prediction tools (Figure 3.4), a central alpha helical domain is present in model 106, spanning residues A326 to R335; exceedingly close to the originally predicted F324 helical termination. This structural feature also corroborates with all PDB templates used during the multi-template modelling stage (Figure 3.4). The second antiparallel β -strand present within the RING 106 model (residues M321 to C325) was absent from both JPred and PsiPred secondary structure predictions (Figure 3.4); but

close to the prediction based on the PDB multi-template consensus (modelled to be positioned H322 to F324). Whilst a first (C-terminal) β -strand was predicted with uncertainty (estimated to encompass residues D313 to D318), the model does include a predecessor antiparallel beta strand, encompassing residues C314 to L317.

3.2.4.2 Model validation using ProCheck and ProSA

Model quality was assessed using a Ramachandran plot produced in ProCheck, indicating 80.5 and 19.5% of residues are within the most favoured and additionally allowed regions, respectively (Appendix 2, Figure A2.1). Whilst exhibiting 90% of residues within the most favoured regions is considered an important feature of a good quality (Morris *et al.*, 1992), the high percentage of residues in most favored and additionally allowed regions; coupled with the absence of residues in generously allowed and disallowed regions, indicate that the model is nonetheless of a high quality.

To assess the overall quality of the model, additional model validation was then conducted in ProSa (Wiedertein and Sippl, 2007). The energy Z-score of the model (106) was calculated as -4.61, whilst PDB: 4AYC (corresponding to the PDB file used for basic modelling) was -4.66. In addition, Z-scores (thus energy distributions) of high quality, X-ray diffraction and NMR-derived structures were plotted with respect to both model and template Z-scores (Appendix 2, Figure A2.2).

Taking in consideration the position of the RING 106 Z-score relative to both high-quality PDB deposited structures (Appendix 2, Figure A2.2) and the 4AYC (RNF8) PDB template used at the basic modelling stage, the quality of the CHFR RING model is exceptionally high.

3.2.4.3 Use of a control model to validate the modelling procedure

To verify the modelling strategy used for CHFR RING modelling and therefore assessment of model accuracy, a general control model was produced. An already known structure of RNF8 was modelled in MODELLER (Sali and Blundell, 1993; Eswarn *et al.*, 2006) using an identical strategy previously outlined via the CHFR RING domain homology modelling process. A BLAST search (Altschul *et al.*, 1997) within the PDB (Abola *et al.*, 1987) deduced 6 potential PDB template files, corresponding to structures with the highest percentage identity to the target (39.13 – 42.22%), medium to high resolution (1.80 to 2.65Å) X-ray crystallographic data sets and low R_{free} values (0.213 – 0.244) (Table 3.6, Figure 3.10).

Table 3.6 RNF8 homologous proteins within the local (PDB derived) database

Name	PDB	Chain	Residue range	% Identity	Resolution	R-free (F)	R-work (W)	F-W
E3 ubiquitin ligase: RNF73 (<i>H. sapiens</i>)	2Y43	A	21-67	39.13	1.80	0.223	0.175	0.048
E3 ubiquitin ligase: RING1B (<i>M. musculus</i>)	2CKL	A	14-60	41.30	2.00	0.213	0.179	0.034
E3 ubiquitin ligase: RING1B (<i>H. sapiens</i>)	3RPG	B	14-60	41.30	2.65	0.243	0.217	0.026
B lymphoma Mo-MLV insertion region: RNF51 (<i>H. sapiens</i>)	2H0D	A	14-60	41.30	2.50	0.244	0.210	0.034
Tripartite motif-containing protein: RNF23 (<i>H. sapiens</i>)	2ECJ	A	14-58	40.00	NA	NA	NA	NA
RING finger protein: RNF141 (<i>H. sapiens</i>)	2ECN	A	14-58	42.22	NA	NA	NA	NA

NA: not applicable, NMR solution structures

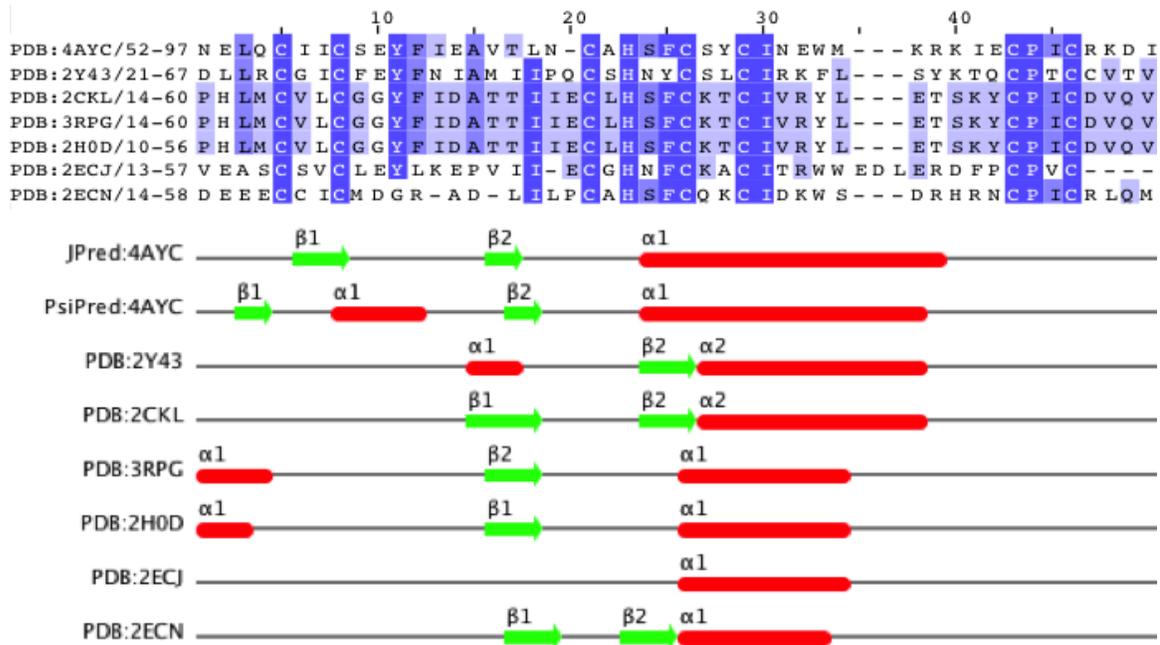


Figure 3.10 Multiple sequence alignment between RNF8 and the local database structures. All homologues within the local PDB database are included within the main alignment. Secondary structural annotations regarding Jpred and PsiPred predictions, in addition to selected PDB structural datasets, are included as alpha helices (red lines) and beta strands (green arrows). Sequence colouration corresponds to identical (dark purple) to partially conserved (light purple) residues between sequences.

Taking into consideration both PDB file (template) secondary structure content; combined with JPred and PsiPred secondary structure predictions, an alpha helix is most likely to encompass the consensus of residues S424 to M382. The significantly variable positioning of beta strands; between PDB structures, PsiPred and JPred secondary structure predictions make it unclear as to whether single or multiple antiparallel beta strands precede the central alpha helix.

Human RNF73 (PDB: 2Y43) was selected as the best template for the single-template stage of homology modelling in MODELLER (Eswarn *et al.*, 2006; Sali and Blundell, 1993); collectively considering the high percentage identity the RNF8 sequence (39.13%), and high resolution of the crystallographic dataset (1.8Å), relative to all other PDB structures. The RNF8 domain boundary was predicted as spanning residues D21 to V67 of the 2Y43 PDB coordinate file; encompassing the entire zinc-binding motif and providing sufficient loop lengths towards the N and C-terminus of the central RING domain and thus stabilizing both zinc ions sufficiently.

For multi-template homology modelling, the top 6 PDB structure datasets (sorted from highest to lowest resolution) were selected, with percentage identities to RNF8 ranging from 39.13 to 42.22%. Alignment between the target (RNF8) and the template structure of RNF73 (PDB: 2Y43) were used to produce 100 models, with best model identified via evaluation of MODELLER-derived DOPE and molpdf scores, representative of the calculated optimized statistical potential (Shen and Sali, 2006) and the sum of all restraints (Sali and Blundell, 1993), respectively. The strategy was repeated for multi-template and loop-refinement modelling methodologies; with the lowest DOPE and molpdf scores tabulated and DOPE/alignment position data plotted (Table 3.7, Figure 3.11) and corresponding to the respective best energy models.

Table 3.7 Evaluation of RNF8 homology models produced in MODELLER

Strategy	Best model (/of)	molpdf	DOPE	molpdf+(-)DOPE
Basic (single template)	85 (/100)	219.304	4546.189	-4081.978
Multi-template	57 (/100)	1689.063	-4524.442	-2835.380
Loop refinement	641 (/1000)	1384.350	-4186.816	-20802.816

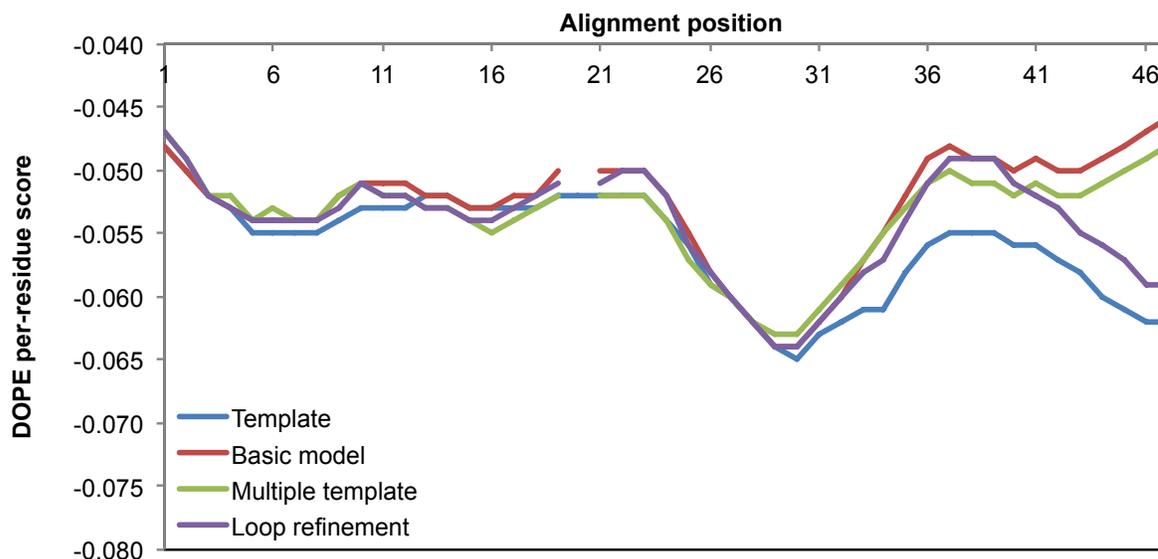


Figure 3.11 DOPE score profile of the RNF8 model and template (PDB: 2Y43) obtained in basic, multi-template and loop refinement modelling strategies in MODELLER. The DOPE profile of the PDB: 2Y43 template is shown in blue, with respective colours representative of the preceding basic, multi template and loop refinement stages. The `evaluate_template4ayc.py` script (Appendix 2, A2.7) was adapted (per model) accordingly.

As identified by the DOPE profile of multi-template model 57 (Table 3.7 and Figure 3.11), residues spanning the RNF8 and PDB: 2Y43 (template) alignment position 31 to 46 (inclusive) exhibited a higher energy profile than the basic model counterpart, indicating requirement of *ab-initio* loop refinement within this region of the model. Multi-template model 57 was therefore selected as the best model for additional loop refinement. Considering none of the first 500 loop models produced featured correct coordination of the C-terminal loop and corresponding residues to the second zinc ion, the number of total models produced was extended to 1000; with approximately 15% exhibiting a slightly lower energy DOPE profile) particularly from alignment position 41 to 47) (Figure 3.11).

Whilst an improvement in DOPE score profile is exhibited within the selected loop-refined models, only one model (641) exhibited correct C-terminal ligand (cysteine or histidine) coordination with the second zinc ion within the RNF8 models (Figure 3.12); exhibiting a slightly lower DOPE profile in comparison to its multi-template model counterpart (number 57).

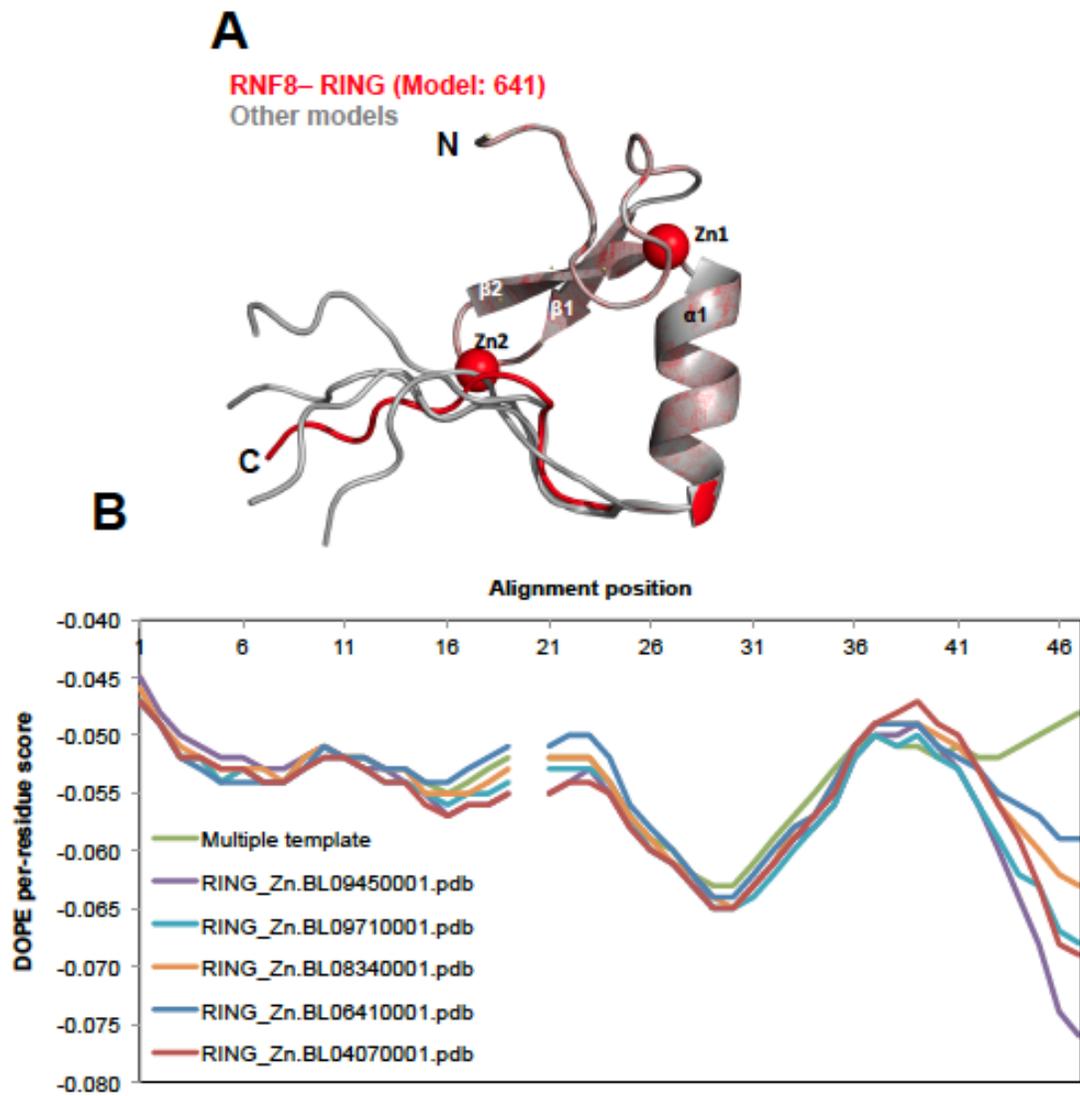


Figure 3.12 (A) Superimposition of the top 5 calculated loop conformations of the RNF8 homology models, refined in MODELLER using the multi-template derived model 57. The best loop-refined model (exhibiting the lowest DOPE profile, spanning alignment position 31 to 47, Figure 15) was identified as Model 641 (Shown in red). Other models with higher molpdf and DOPE scores are shown in grey. Zinc ions (corresponding to model 641) are shown as red spheres. (B) DOPE score profile of the CHFR RING models loop-refined between alignment positions 31 to 46 in MODELLER. The `Loop_refine.py` and `evaluate_template4ayc.py` scripts (Appendix 2, A2.8 and A2.7, respectively) were adapted accordingly.

To observe secondary structure content and correct zinc ion coordination between zincs (1 and 2) and respective (cysteine and histidine) ligands, RNF8 RING model 85 was viewed as a cartoon in PyMol and aligned to the template structure (used in basic modelling) of RNF8 (PDB: 2Y43) (Figure 3.13).

4AYC-RING (Model: 85)

PDB: 2Y43 (RNF73)

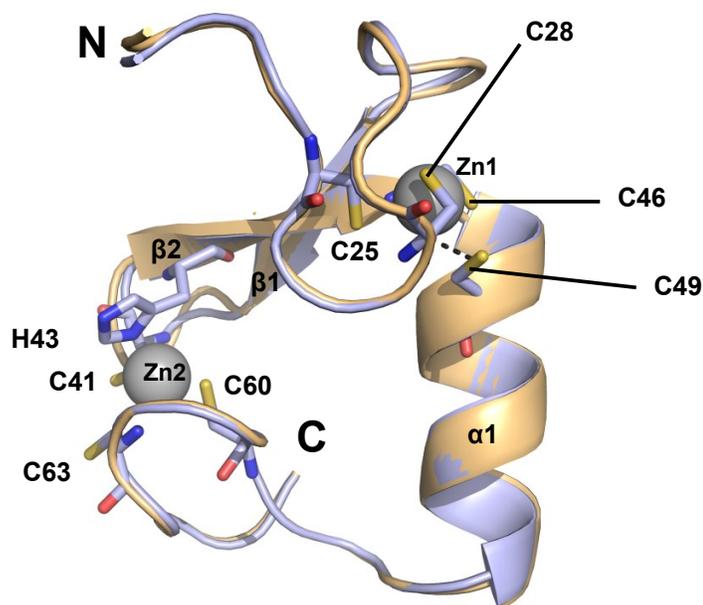


Figure 3.13 Cartoon alignment between PDB: 2Y43 (basic modelling template) and model 85, produced by homology modelling using MODELLER. Alignment of both model and correct structure indicate absolutely identical positioning of both antiparallel β strands (1 and 2). Residues coordinating to zinc ion metal centres 1 (Zn1) and 2 (Zn2) are shown as sticks, whilst inc ions are shown as grey spheres.

Model quality of the control (641) in comparison to the deposited RNF8 structure was assessed using a Ramachandran plot produced in ProCheck, indicating 88.4 and 11.6% of residues are within the most favoured and additionally allowed regions of the model, respectively – a noticeable improvement, in comparison to the actual deposited structure (Appendix 2, Figure A2.3).

To assess the overall quality of the model with respect to the deposited PDB structure further, additional model validation was then conducted in ProSa (Wiedertein and Sippl, 2007). The energy Z-score of the model (641) was calculated as -4.69, whilst PDB: 4AYC was -4.66. In addition, Z-scores (thus energy distributions) of high quality, X-ray diffraction and NMR-derived structures were plotted with respect to both model and template Z-scores (Appendix 2, Figure A2.4). Taking in consideration the position of the RNF8 model 641 Z-score relative to both high-quality PDB deposited structures (Appendix 2, A2.4) and the actual PDB structure (PDB: 4AYC), the validity and quality of the RNF8 model is exceptionally high.

To assess whether the secondary structural features of the RNF8 control model correspond to the PDB deposited structure, an additional alignment of the cartoon 641 model and actual structure was made in PyMol (Figure 3.14).

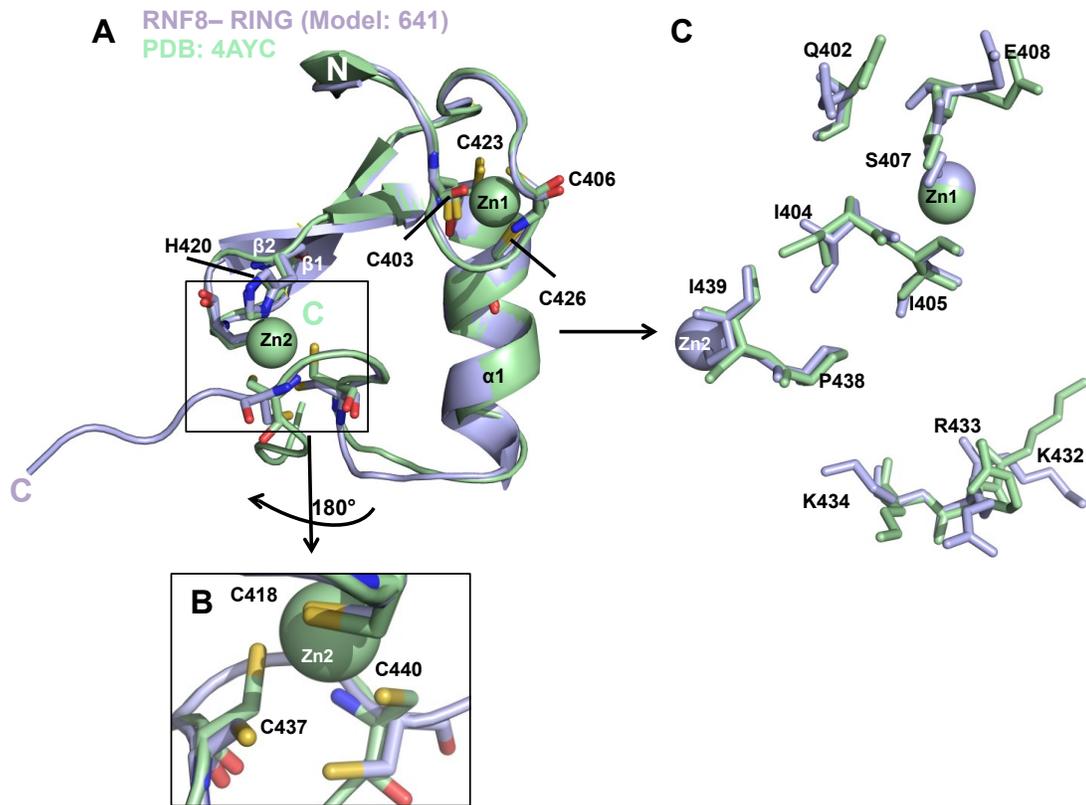


Figure 3.14 Alignment between RNF8 RING model (641) and PDB deposited structure (4AYC). (A) Alignment between both cartoons indicates that size of both model and PDB file beta strand 1 (residues A413 to L416) are identical, whilst the second antiparallel strand of the model (S421 to C423) is slightly shorter than the deposited structure counterpart (A419 to C423). In addition, the central alpha helix within the RNF8 RING domain is slightly shorter within the model (S424 to M431), in comparison to the deposited structure, whereby the alpha helix is extended as far as R433. Within the model, the first N-terminal alpha helix is missing, unsurprising considering there is an insufficient hydrophobic- hydrophilic residue periodicity that would otherwise facilitate MODELLER prediction of this secondary structural element. All other loops within the model directly correspond to the PDB equivalent, with the exception of C441 to I444 of the N-terminus, indicating a potential source of error. (B) Looking more closely at two of the three cysteine residues coordinated to the second zinc ion (C418, C440 and C437), there appears to be a change in protein conformation, in comparison to the deposited structure. C437 and C440 appear in different orientations in comparison to the PDB structure. Nonetheless, correct coordination with the second zinc ion centre (ZN2) is observed. (C) Residues involved in RNF8 recognition of UbCH13 were also mapped between structures (shown here as sticks, Campbell *et al.*, 2012); with incorrect orientation, relative to the deposited structure, in of some residues observed. Zinc ions are shown as spheres.

Root mean square deviations (RMSD) between the model and PDB counterpart (PDB: 4AYC, calculated in PyMol) C α -C α backbone and entire structure was calculated as 0.468 and 0.573Å, respectively, indicating a high level of accuracy is present within the model.

Taken together, the stereochemistry and quality (assessed by ProSa) indicate model 641 of RNF8 is a high quality model, with some individual residue conformations and N-terminal loop deviations from the PDB (4AYC) deposited structure. This indicates that the accurate modelling strategy employed in predicting the RNF8 RING domain structure is concurrent with that implemented to predict the CHFR RING domain structure.

3.2.4.4 Iterative modelling and extension of the CHFR RING domain

In order to predict residues at the CHFR interface with the E2 heterodimer Ubch5a: Ub, the final CHFR RING model (10) cartoon was aligned (in PyMol) with the deposited complex of (a fused RING dimer of) RNF4: Ubch5a ~ ubiquitin (PDB: 4AP4, Plechanovova *et al.*, 2012). In addition, a multiple sequence alignment and secondary structural annotations were cross-referenced with the aligned cartoons (Figure 3.15).

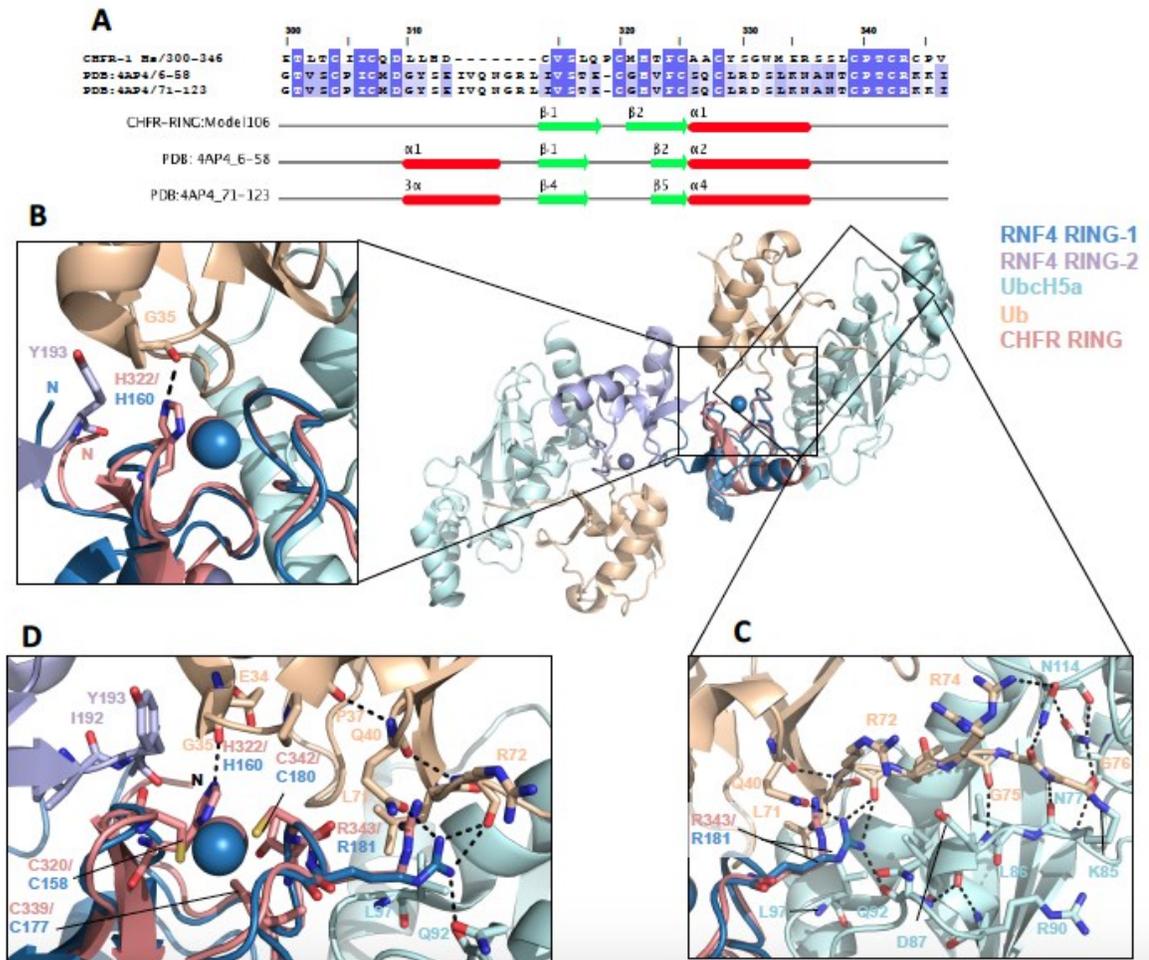


Figure 3.15 Alignment between the final CHFR RING model (106) and the E3 ubiquitin ligase homologue RNF4 (PDB: 4AP4, Plechanovova *et al.*, 2012). (A) A multiple sequence alignment between a fused RNF4 dimer and CHFR RING domain corresponds to a 34.04% sequence identity (Table 3.3), with identical residues coloured in purple and semi-conserved residues (between the two individual RNF4 RING domains) coloured in light purple. The amino acid numbering used corresponds to the CHFR RING-domain (Isoform 1). RNF4 is ordinarily monomeric; however, prior to crystallization of the complex, a fused dimer of the domain was produced by Plechanovova *et al.* (2012); containing two separate zinc binding RING domains per individual E3 ubiquitin ligase monomer. This is in contrast to the single RING domain predicted within CHFR. Both structures share a significant amount of secondary structure content, including two antiparallel beta strands (C304 to C307 and M326 to C325, CHFR RING, green arrows) and a central alpha helix (A326 to R335, CHFR RING, red line). However, CHFR Model 106 lacks an additional alpha helical domain (3), as presented within the two individual RNF4 RING domains. (B) The cartoon of the CHFR RING model was then aligned with RNF4 and key residues at the various interfaces (CHFR RING: Ubch5A: Ub and Ubch5A:Ub) extrapolated from the original RNF4:Ubch5A:Ub crystal structure (in a 1:2:2 ratio) and shown as sticks (Plechanovova *et al.*, 2012). A single CHFR RING domain is shown aligned with a single RNF4 RING domain (RING-1 in dark blue). (B) CHFR RING side chain residue H322 (positioned at zinc metal ion centre 2) appears to make a stacking interaction with the Ub main chain E34 and G35 residues; corroborating with the RNF4 H193 residue counterpart (Plechanovova *et al.*, 2012).

Figure 3.15 Alignment between the final CHFR RING model (106) and the E3 ubiquitin ligase homologue RNF4 (PDB: 4AP4, Plechanovova *et al.*, 2012). (Legend continued...).

However, since this interaction is located at the CHFR N-terminus, the RING domain model terminates and therefore the RNF4 side-chain Y193 residue from RING-2 (in purple, stacking also with Ub G35 and H193) and has no CHFR RING equivalent modelled. Dimer formation is therefore required for RNF4 interaction with Ub (Plechanovova *et al.*, 2012). (C) As previously described (Plechanovova *et al.*, 2012), ubiquitin forms a considerable interface with Ubch5a, including 4 salt bridges and 15 hydrogen bonds across 1,800Å². Both T341 and R343 of the CHFR RING domain forms polar contacts within the Ubch5a:Ub interface; to residues L71, R72 and L8 to K11 of Ub, and Q92 and L97 of Ubch5a. These residues correspond to T179 and R181 of RNF4 (Plechanovova *et al.*, 2012). (D) Looking at the Ub hydrophobic cluster in more detail (L8, T9 and L71), an interface with Ubch5a (A96 and L97) can be observed; with CHFR RING R343 (I305 and P340, not shown) projecting into the Ubch5a: Ub interface, with similar interactions observed in RNF4 (R181, P137 and P178, respectively) (Plechanovova *et al.*, 2012).

Taken together, the three-molecule interface interactions previously described within the RNF4: Ubch5a: Ub complex via the co-crystal structure obtained (Plechanovova *et al.*, 2012) is very similar to that observed within the RING CHFR model. However, both RING domains within the single RNF4 fusion (RING-1 and RING-2) are essential for its ubiquitin ligase activities mediated by Ubch5a; with both domains also making contact with the Ubch5a E2 ubiquitin conjugating enzyme (Plechanovova *et al.*, 2012). In addition, Plechanovova *et al.* (2012) previously identified the conservation of consecutive aromatic hydrophobic – hydrophobic residues within dimeric RING-type E3 ubiquitin ligases via a multiple sequence alignment. However, the CHFR RING domain model terminated at the N-terminus, prior to additional modelling of a potential interaction employing any residue corresponding to Y193 in the RNF4 RING-2 domain (RING-2, Figure 3.15, Panels B and D).

This therefore lead to the proposition that an additional iterative modelling step could be used, in an attempt to extend the original CHFR RING model N-terminus and therefore include any residues with similar properties (to RNF4 RING-2 residue Y193). This may prove informative in both predicting an additional interaction between the CHFR RING domain and ubiquitin, whilst predicting the resulting stoichiometry underpinning RING: Ubch5a: Ub complex formation and ubiquitination mechanism thereof.

To integrate the additional Y361 residue of the CHFR RING domain model within the predicted UbCH5a: Ub interaction, the original RING-106 model was extended to include Y362 and L363 (corresponding to Y193 and I194 of RNF4, respectively). A basic model was produced using the RNF4 structure corresponding to PDB dataset 4AP4 (Plechanovova *et al.*, 2012) using MODELLER (Eswarn *et al.*, 2006; Sali and Blundell, 1993). A multi-template model was then produced, using the original RING CHFR model 106 and PDB: 4AP4 (Plechanovova *et al.*, 2012) as a template (to ‘anchor’ the loop and Y193 residue in the correct position during the modelling process). A multiple sequence alignment between RNF4 (PDB: 4AP4) and CHFR-RING model indicate conservation of a Y (hydrophobic) residues pattern; with a 30.64% sequence identity calculated (via the SIAS server, (SIAS, 2016) between RNF4 and CHFR RING domain within the new alignment (Figure 3.15).

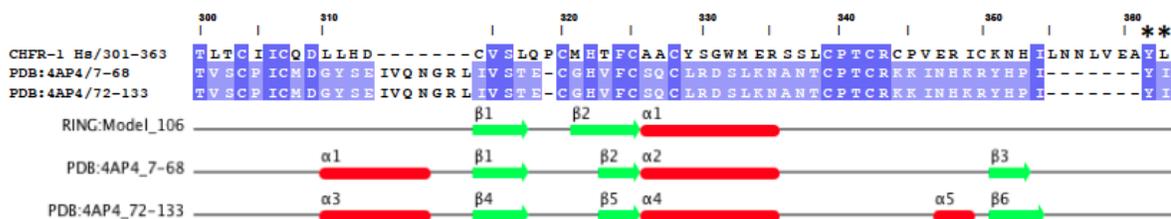


Figure 3.16 Alignment between Model 106 of the CHFR RING domain with RNF 4 (PDB: 4AP4, Plechanovova *et al.*, 2012). The alignment indicates positioning of $\alpha 1$ and $\alpha 3$ (corresponding to RING-1 and RING-2 domains in the fused RNF4 dimer), responsible for RNF4 dimer formation and absent within the CHFR RING model and are instead represented by a gap (between CHFR RING residues D313 and C314). Conserved tyrosine and hydrophobic (non-aromatic) residues between RNF4 and CHFR RING are indicated by the asterisks (*). Red lines depict alpha helices, whilst green arrows represent beta strands. Amino acid residues denoted correspond to CHFR RING isoform 1, with dark purple shading representing identical residues and lighter purple residues identical within the RNF4 fusion peptide.

The best models were identified via evaluation of MODELLER-derived DOPE and molpdf scores, representative of the calculated optimized statistical potential (Shen and Sali, 2006) and the sum of all restraints (Sali and Blundell, 1993), respectively. The strategy was repeated for multi-template and loop-refinement modelling methodologies; with the lowest DOPE and molpdf scores tabulated and DOPE/alignment position data plotted (Table 3.8, Figure 27) and corresponding to the respective best energy models.

Table 3.8 Evaluation of CHFR RING domain homology models produced in MODELLER

Strategy	Best model (/of)	molpdf	DOPE	molpdf+(-)DOPE
Basic (single template)	68 (/100)	686.334	-6101.155	-5414.821
Multi-template	37 (/100)*	1336.542	-6274.873	-4938.330
Loop refinement	13 (/1000)*	29.936	-5674.281	-5644.345

* N.B: Initially evaluated as the 'best' models, see next section for more information).

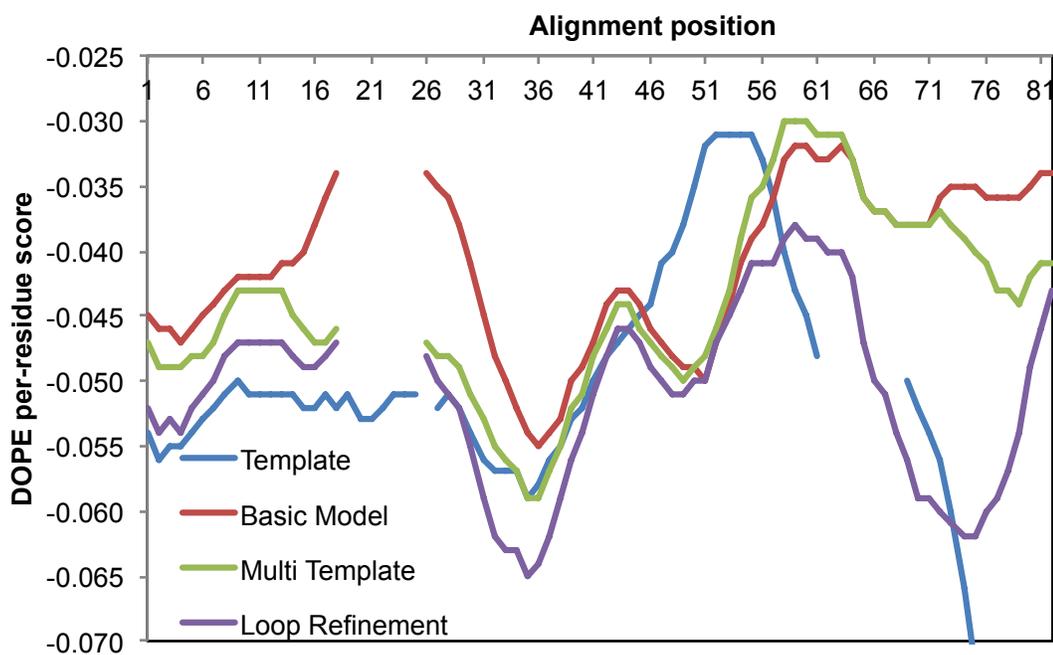


Figure 3.17 DOPE score profile of the CHFR RING extended model and template (PDB: 4AYC) obtained in basic, multi-template and loop refinement modelling strategies in MODELLER. The DOPE profile of the PDB: 4AP4 template is shown in blue, with respective colours representative of the preceding basic, multi template and loop refinement stages. The `evaluate_template4ayc.py` script (Appendix 2, A2.7) was adapted (per model) accordingly.

As identified by the DOPE profile of multi-template model 37 (Table 3.8 and Figure 3.17), residues spanning the RING model 106 and PDB: 4AP4 (template) alignment position 58 to 82 (inclusive) exhibited a higher energy profile than the basic model counterpart, indicating requirement of *ab-initio* loop refinement within this region of the model. Multi-template model 58 was therefore selected as the best model for additional loop refinement. From 500 models produced by the loop

refinement strategy, ~12% exhibited correct zinc ion (2) and ligand (CHCC) coordination, and were therefore selected for additional evaluation; with DOPE and molpdf scores tabulated and DOPE/ alignment position data plotted (Figure 3.18) and corresponding to the best energy models.

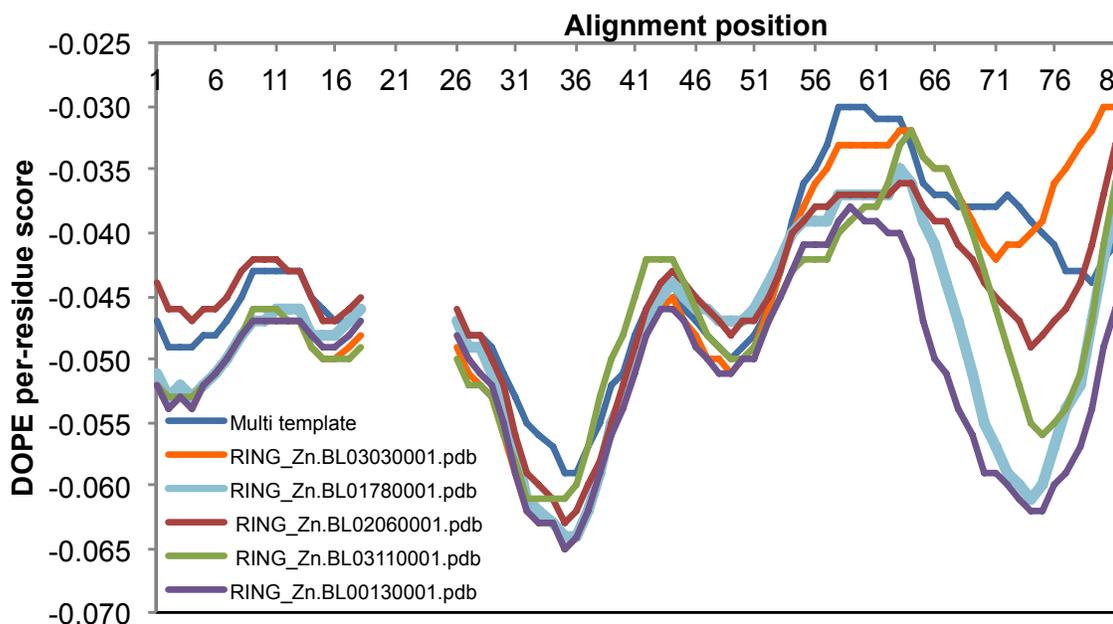


Figure 3.18 DOPE score profile of the extended CHFR RING domain models loop-refined between alignment positions 58 to 82 in MODELLER. The `Loop_refine.py` and `evaluate_template4ayc.py` scripts (Appendix 2, A2.8 and A2.7, respectively) were adapted accordingly.

Whilst an improvement in DOPE score profile is exhibited within the selected loop-refined models, no models exhibited correct C-terminal ligand (cysteine or histidine) coordination with the second zinc ion (Figure 3.19).

CHFR-RING extended (Model: 13)

Other models

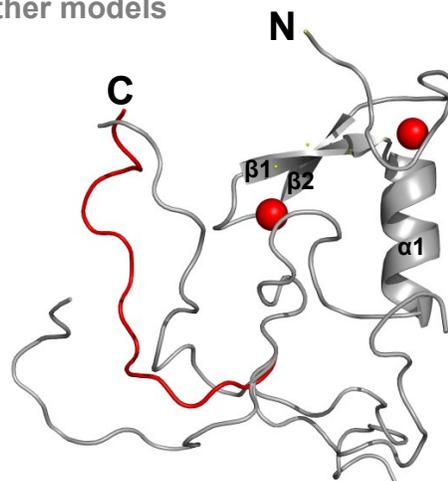


Figure 3.19 Superimposition of the top 5 calculated loop conformations of the extended CHFR RING domain, refined in MODELLER using the multi-template derived model 37. The best loop-refined model (exhibiting the lowest DOPE profile, spanning alignment position 58 to 82, Figure 29) was identified as Model 13 (Shown in red). Other models with higher molpdf and DOPE scores are shown in grey. Zinc ions (corresponding to model 13) are shown as red spheres.

Despite the improved DOPE profile corresponding to model 13 (Figure 3.18), incorrect coordination of the second zinc metal ion centre with respective (ligand) residues (CHCC) was observed (Figure 3.19). In addition, all remaining models did not include correct loop conformation within the CHFR RING N-terminus (data not shown). Therefore, the previous models obtained via multi-template modelling using the original CHFR RING model 106 and RNF4 (PDB: 4AP4) were re-evaluated in more detail. The top-5 multi-template models with the lowest molpdf +(-) DOPE score were therefore evaluated in more detail, with initial visualisation of zinc coordinating metal ion centres 1 and 2 indicating correct loop conformation and ligand interactions were present in all models (Table 3.9). In addition, Ramachandran plots were produced of all five models to assess individual stereochemical quality (Appendix 2, Figure 2.5).

Table 3.9 Summary of MODELLER evaluation data and PROCHECK stereochemical quality assessment of extended, multi-template CHFR RING domain models

Rank	Model	DOPE	Molpdf	Molpdf + (-) DOPE	Residues in most favored regions (%)	Residues in additional allowed regions (%)	Residues in generously allowed regions (%)	Residues in disallowed regions (%)
1	37	-6274.873	1336.542	-4938.330	54 (7.94)	12 (17.6)	1 (1.5)	1 (1.1)
2	54	-6335.328	1424.975	-4910.353	58 (85.3)	9 (13.2)	1 (1.1)	0 (0.0)
3	38	-6263.335	1359.471	-4903.864	56 (82.4)	10 (14.7)	0 (0.0)	2 (2.9)
4	23	-6339.474	1436.273	-4903.201	53 (77.9)	12 (17.6)	2 (2.9)	1 (1.5)
5	10	-6258.062	1369.824	-4888.237	58 (85.0)	10 (14.7)	0 (0.0)	0 (0.0)

Since model 10 contains no residues within generously allowed and disallowed regions, it was considered the best model of the extended CHFR RING domain. A cartoon comparison between RNF4 (PDB: 4AP4) and the (extended) CHFR RING model 10 indicates regions of both secondary structure insertion and deletion between the E3 ubiquitin ligase homologues (Figure 3.20).

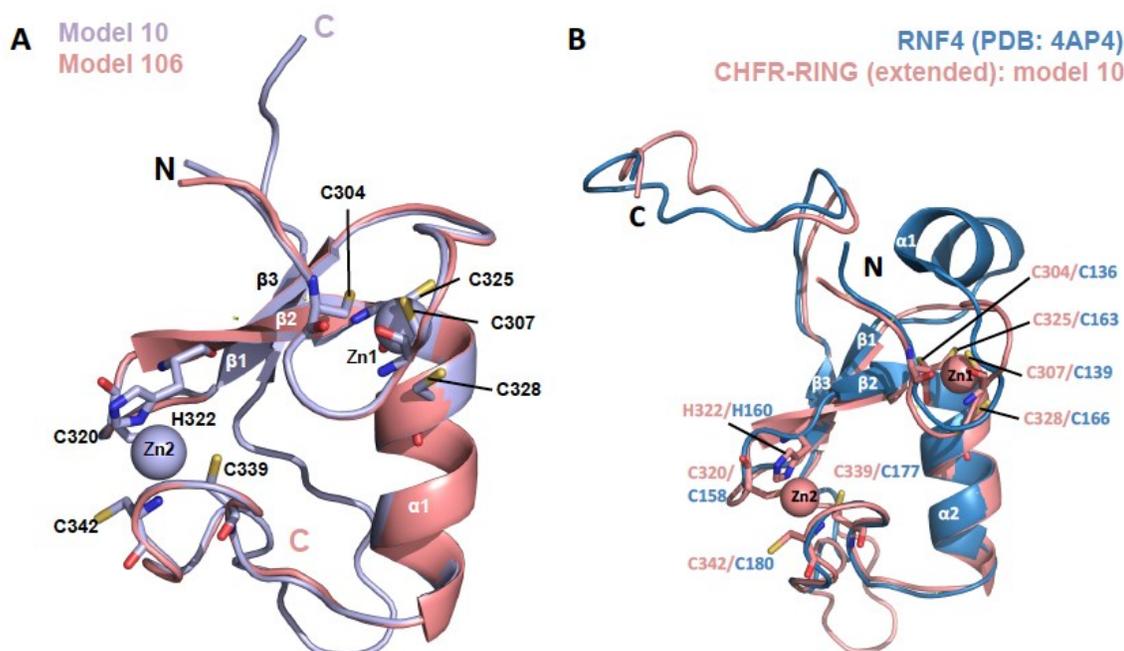


Figure 3.20 Cartoon alignments between (A) CHFR RING models 106 and 10 and (B) template (RING model 106 and RNF4) and RNF4 (PDB: 4AP4), produced by homology modelling using MODELLER. (A) Positioning of loops and secondary structural elements are identical to Model 106, unsurprising considering the structural coordinates were included in the multi-template modelling. However, an additional antiparallel beta strand (3) is predicted in model 10 (L358 to E360). (B) Alignment of both model and template indicate identical positioning of both antiparallel β strands (1 and 2); with the second CHFR β strand slightly longer than the template equivalent. The N-terminal α -helix of the 4AYC PDB template is absent from the CHFR RING model. Alpha helical numbering corresponds to the 4AYC PDB template. Alpha helix 1 corresponds to the insertion observed specifically in RNF4 (spanning residues G142 to N149, Figure 26) and absent from the CHFR RING model counterpart. The additional loop region within the CHFR RING model (residues L350 to E360) corresponds to an insertion absent from the RNF4 sequence (Figure 26). Residues coordinating to zinc metal ion centres 1 and 2 (shown as spheres) are coordinated by their respective cysteine and histidine residues, shown as sticks.

A Ramachandran plot of the new model indicates that the percentage of residues present in allowed and additional allowed regions of model 10 is higher than its model 106 counterpart (Appendix 2, Figure A2.6).

Whilst exhibiting 90% of residues within the most favoured regions is considered an important feature of a good quality (Morris *et al.*, 1992), the high percentage of residues in most favored and additionally allowed regions; coupled with the absence of residues in generously allowed and disallowed regions, indicate that model 10 is nonetheless of a high quality in comparison to its predecessor, model 106. To assess the overall quality of the model, additional model validation was then conducted in ProSa (Wiedertein and Sippl, 2007) (Appendix 2, Figure A2.7).

The energy Z-score of model 106 was calculated as -4.61, whilst model 10 was -3.24. Taking in consideration the position of the RING 106 and RING 10 Z-score relative to both high-quality PDB deposited structures (Appendix 2, Figure A2.7), the validity of both CHFR RING models is comparable to that of known (PDB deposited) structures.

Alignment of CHFR RING models (106 and 10) with respect to RNF8 (Campbell *et al.*, 2012) indicate that the loops within the new extended model 10 corroborate with RNF8 and RING model 106 (Figure 3.21). This strongly suggests that residues present at the RING CHFR model 10: Ubc13: Mms2 interface are identical to the RING model 106 counterpart (Figure 3.21).

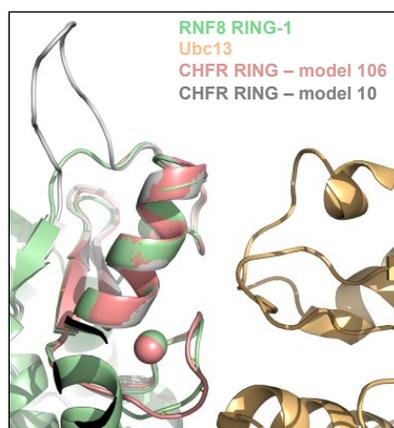


Figure 3.21 Alignment between RNF8 (PDB: 4ORH, Campbell *et al.*, 2012) RING model 106 and RING model 10. Structures are shown as cartoons, with zinc ions represented by spheres.

With the new RING 10 extended CHFR model, its interaction with UbcH13 was then reassessed (Figure 3.22).

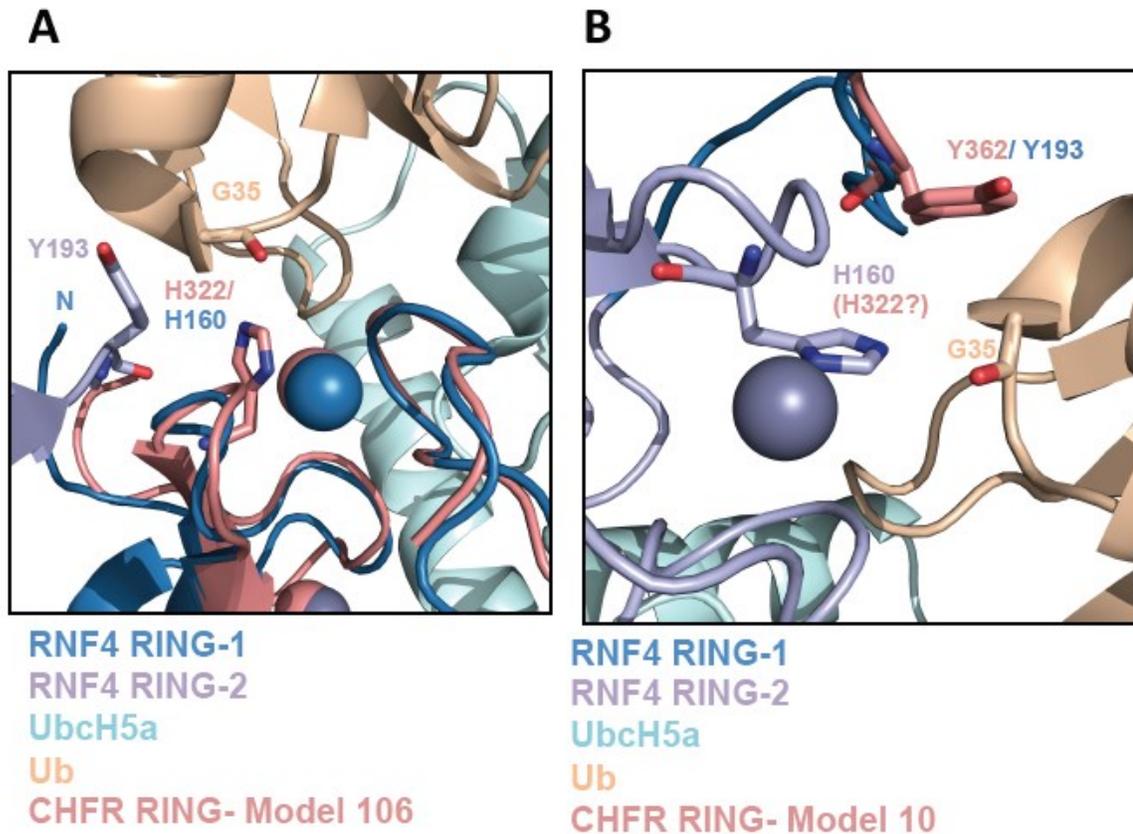


Figure 3.22 Re-evaluation of the CHFR RING: Ubiquitin stacking interaction of (A) model 106, in comparison to (B) model 10. (A) CHFR RING Model 106 was aligned with the RNF4 second zinc RING-2, with an interaction formed between Ub G35 and H322. Any preceding residues within the domain were not modelled, since the N-terminus of the domain excluded any other preceding residues. (B) The new model (RING-10) was produced using a multi-template modelling strategy, inclusive of the RING 106 model and RNF4 (PDB: 4AP4). The Y363 residue now extends into the stacking interaction formed by RNF4 RING-2 zinc binding domain; including Ub G35 (from another interaction) and H160. According to both models, CHFR contains one zinc binding domain suggesting that a H322 residue from second RING domain would be required to complete the stacking interaction and corroborate with the RNF4 findings described by Plechanovova *et al* (2012). Key residues are shown as sticks, whilst zinc ion metal centres are shown as spheres.

A similar modelled stacking interaction can also be observed in other RING E3: E2 ~ Ubiquitin complexes; whereby the CHFR RING model 10 was also superimposed (using Pymol) with a fused RNF4 dimer (*R. norvegicus*) in complex with Ubc13 and polyubiquitin C (PDB: 5AIU, Branigan *et al.*, 2015) (Figure 3.23).

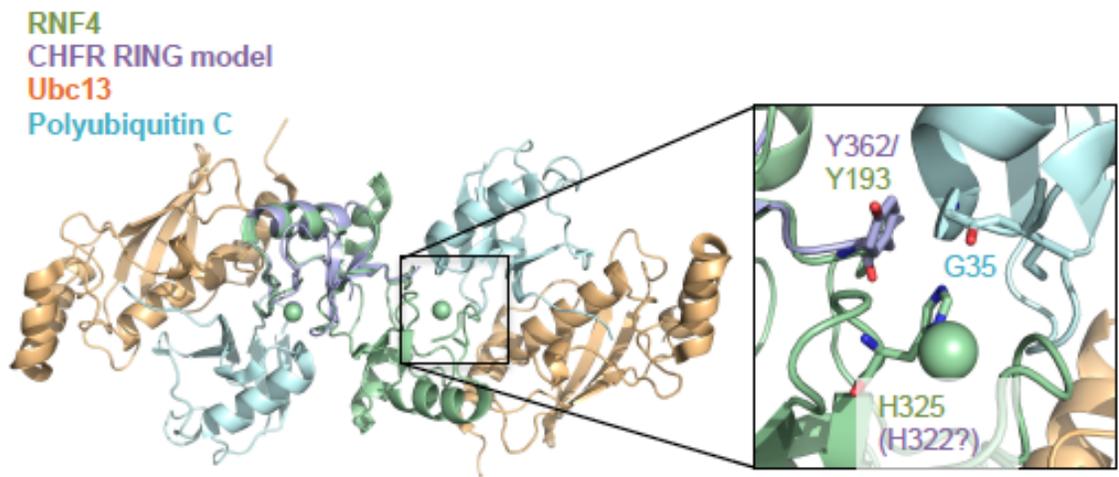


Figure 3.23 Re-evaluation of a potential stacking interaction of CHFR RING model 10 and ubiquitin within the RING: Ubc13 ~ ubiquitin complex. The extended CHFR RING Model (10) was superimposed on RNF4, within the crystal structure of an RNF4 (fused dimer) with polyubiquitin C and Ubc13 (Branigan *et al.*, 2015; PDB: 5AIU). Stacking between CHFR RING -1 Y362 with H322 of a second CHFR RING centre (RING-2) (or Y193 and H325 within the E3 RNF4 homologue, respectively) may also interact with G35 of ubiquitin, within the proposed model. Key residues are shown as sticks and zinc metal ions are shown as spheres.

3.2.5 Molecular dynamic simulations and evaluation of protein model stability

MD studies are important in understanding both energies and molecular geometries of molecules; not as static, lone entities but instead interacting with environmental features and potentially each other (Schlink, 2010). Therefore, gaining a greater understanding of the dynamic motions underpinning molecules can facilitate understanding the associations between known protein sequence, structures and function by explaining the thermoaccessibility within specific system states (Schlink, 2010). MD simulations of the CHFR RING (model 10) were used to provide information and additional comparisons, regarding the dynamic and structural properties (Schlink *et al.*, 2010) of the CHFR RING domain.

A MD simulation process was implemented to explore the stability of the second (extended) CHFR RING model 10 in the presence of metal zinc ions. During system preparation, 8923 TIP3P (Jorgensen and Tiradorives, 1988) water

molecules were added to the CHFR RING protein model 10, and then neutralized by the addition of 6 NA ions (Figure 3.24).

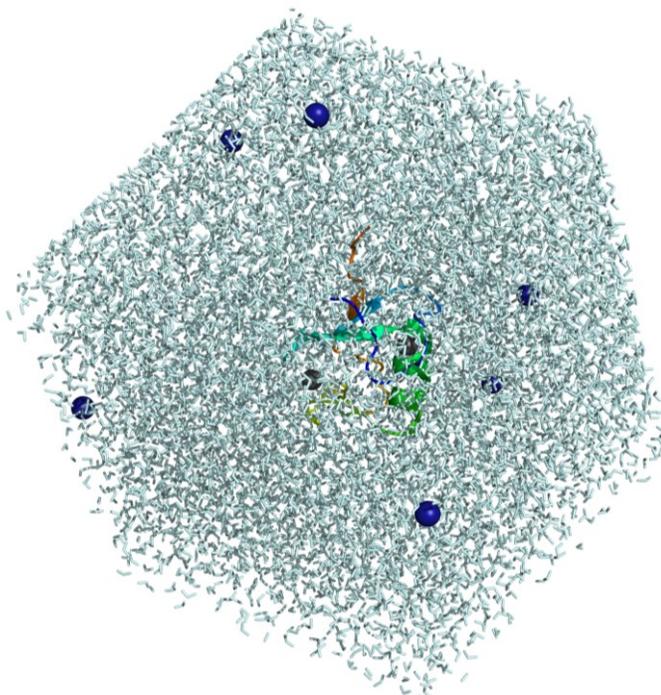


Figure 3.24 CHFR RING model 10 (with zinc ions) solvated and neutralized within a rhombic dodecahedron triclinic unit cell (solvated and neutralized in GROMACS; Van der Spoel *et al.*, 2005). The 6 neutralizing NA ions are shown as navy spheres, whilst the cartoon CHFR molecule is depicted with a rainbow colouring, specifying the N (blue) and C (red) termini. Zinc metal ion centres are shown as grey spheres, with 8923 TIP3P (Jorgensen and Tiradorives, 1988) water molecules within the rhombic dodecahedron unit cell (283.70 nm^3) shown as light blue sticks.

The electroneutral and solvated system was equilibrated prior to the dynamic phase of the simulation. The system was subjected to energy minimization using the steepest descent algorithm (Apol *et al.*, 2010), with a maximum of 50,000 steps used to reach a maximum force (F_{max}) of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$, which was deemed the most appropriate gradient to produce a relaxed CHFR RING model (with zinc ions) (Appendix 2, Figure A2.8). Prior to beginning dynamic studies using the CHFR RING 10 model, the temperature, pressure and density of the system was equilibrated (Appendix 2, Figure A2.9).

The average density of the system using the TIP3P (Jorgensen and Tiradorives, 1988) water model was $1001.83 \text{ Kg m}^{-3}$, very close to the actual density of the

water model at 300K ($1002.00 \pm 1.00 \text{ Kg m}^{-3}$; Mahoney and Jorgensen, 2000) (Appendix 2, Figure A2.9); indicative that the density of the system is well-equilibrated. In addition, the plotted 10-point running averages of temperature, pressure and density (with averages of 299.997 K, 3.57597 bar and $1001.83 \text{ Kg m}^{-3}$, respectively) do not strongly fluctuate about 300K, 0 and 1000 Kg m^{-3} , respectively (Appendix 2, Figure 2.9). This indicates that all three trajectories within the initial 500 ps equilibrium run have stabilized the system for proceeding MD simulation experiments.

Conformation stability of the CHFR RING domain, inclusive of zinc ions, was investigated by conducting 70 ns MD simulations (in triplicate) using the final homology model (10), with zinc ions. Analysis conducted included: root-mean squared deviations (RMSDs) of the protein backbone and radius of gyration (R_g) (Figure 3.25; Panels A and B, respectively).

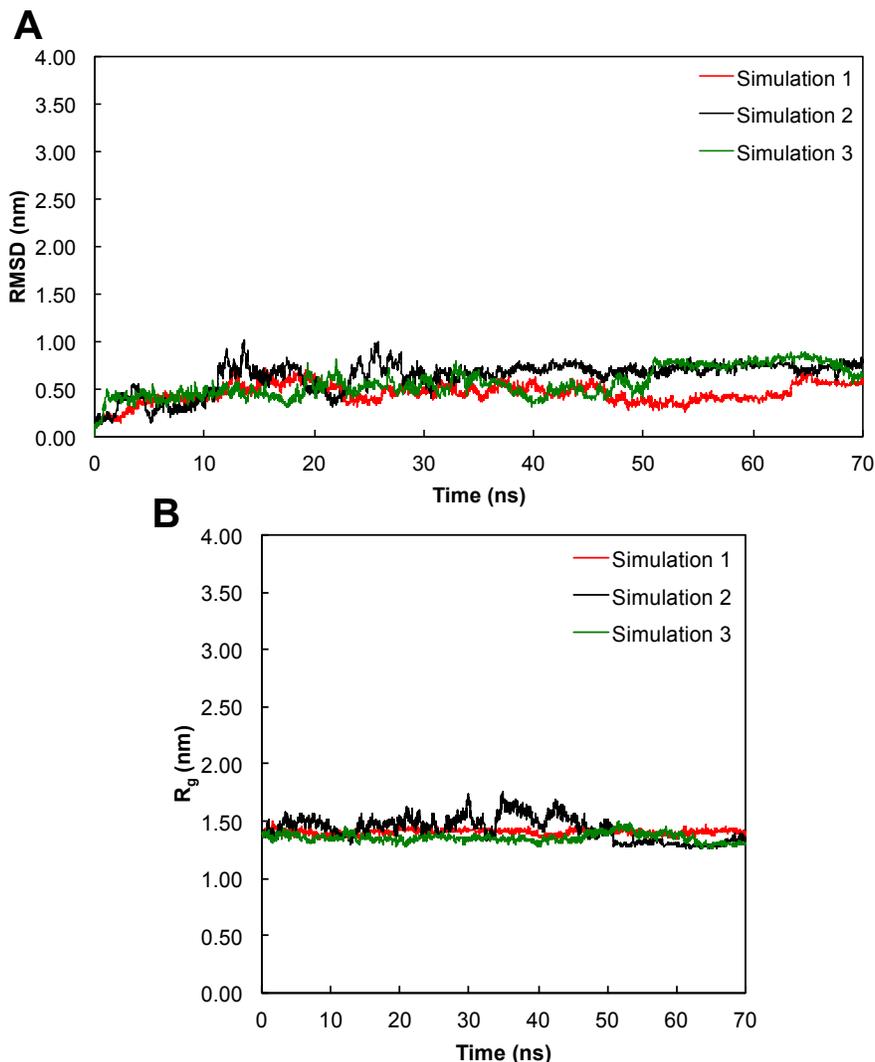


Figure 3.25 Trajectories for CHFR RING (model 10) simulations with zinc. (A) Root mean squared deviations (RMSD) with respect to the CHFR RING model 10 backbone were calculated in GROMACS. The CHFR RING model (within a rhombic dodecahedron) was initially solvated via the TIP3P water model and neutralized. Production runs were carried out in 2,500,000 steps, with constant pressure, isotropic scaling and 2.0 ps relaxation time. Coordinates, velocities and energies were saved every 10.0ps for 5000 steps. (B) Radius of gyration (R_g) of the CHFR RING model was also calculated during the production run, implemented in 2,500,000 steps; with constant pressure, isotropic scaling and 2.0ps relaxation time. Coordinates, velocities and energies were saved every 10.0ps for 5000 steps. All simulations were performed in triplicate, as indicated by the red, black and green traces.

Along the 70 ns trajectory, RMSD values of atomic coordinates in each frame with respect to the original CHFR RING backbone model fluctuated within all three simulations (1 to 3) and around the mean values of 0.46 ± 0.10 , 0.63 ± 0.16 and 0.57 ± 0.15 nm (Figure 3.25, Panel A). With average fluctuation of 0.55 ± 0.14 nm the entire 70 ns trajectory, all three simulations (1 to 3) stabilize after 50 ns and

around the mean RMSD value of 0.65 ± 0.07 nm; indicative of the CHFR RING model stability in water.

The radius of gyration (R_g) of the CHFR RING model with zinc ions during the 70 ns production run (Figure 3.25, Panel B) remains considerably stable for 2 of the 3 simulations (1 and 3, indicated by red and green traces, respectively). However, all three stabilize (inclusive of simulation 2, indicated by the black trace) after 50 ns with an average R_g of 1.36 ± 0.01 nm, indicative of the CHFR RING model's stability and good maintenance of fold across all three independent simulations in water.

Root Mean Square Fluctuations (RMSF) are indicative of the standard deviations present of atomic positions within the trajectory, fitted to a reference frame. Average RMSF values (and therefore local mobility of residues within the simulations) were calculated for each (whole) residue within the RING CHFR model 10 across all 3 simulations (Figure 3.26).

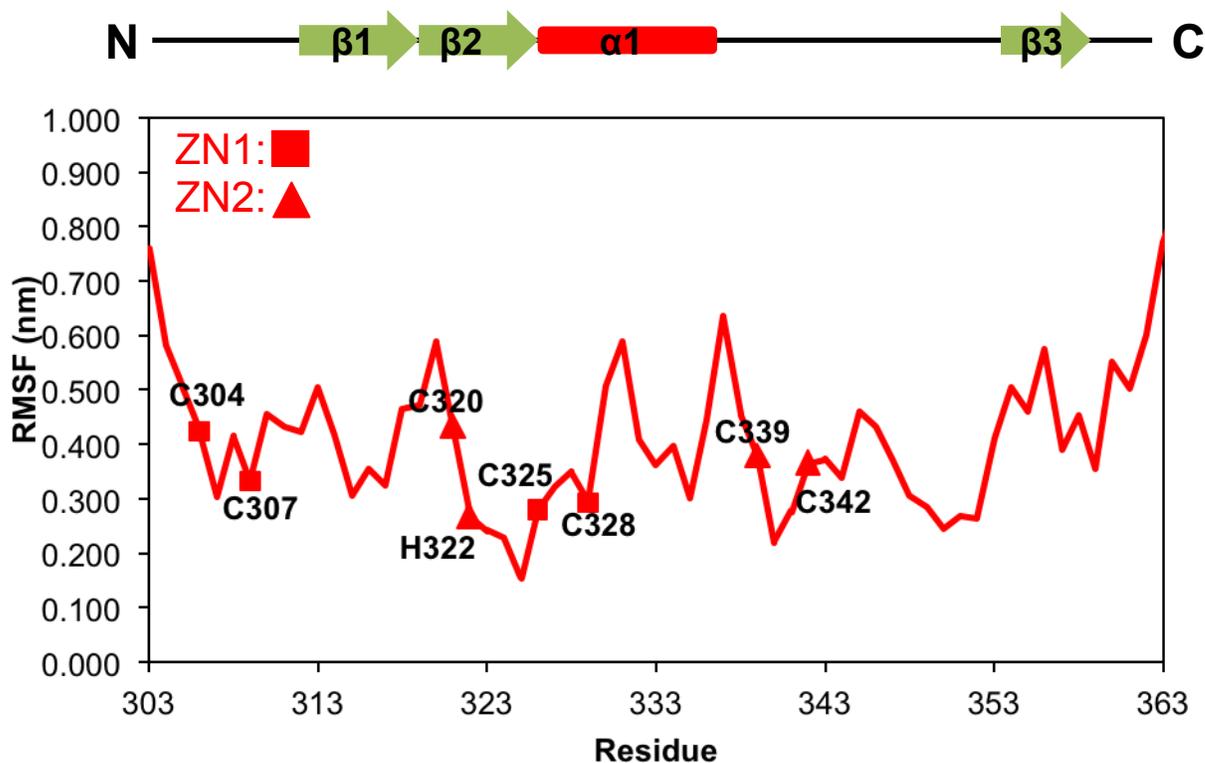


Figure 3.26 Root mean squared fluctuations (RMSFs) with respect to the CHFR RING model 10 structures, with RMSF values calculated for individual (whole) residues. Modelled CHFR RING model 10 secondary structural elements are also included above the plot, corresponding the relative positions of beta strands 1, 2 and 3 (green arrows) and central alpha helix (red line). Squares and triangles represent residues co-ordinated to zinc ions 1 and 2 (ZN1, -2) respectively.

Standard deviations of CHFR RING model 10 residues, fitted within the trajectory, are higher within loop regions of the model, in comparison to secondary structural counterparts (Figure 3.26). Standard deviation of RMSF values within the CHFR RING N-terminal loop region (encompassing residues T303 to C314 and ranging from 0.759 to 0.417 nm, respectively) at ± 0.119 nm is higher in comparison to the ± 0.001 nm equivalent across the central α -helix (spanning residues A326 to R335) (Figure 3.26). This indicates a higher local mobility and flexibility (thus fluctuations in RMSF values) of residues within loop regions, in comparison to secondary structure counterparts.

3.3 DISCUSSION

3.3.1 Secondary structure prediction via PBD structure content proves more informative, in comparison to JPred and PSI-Pred secondary structure prediction tools

In the absence of a CHFR RING domain experimental structure, a homology modelling was produced, whereby closely related RING domain containing homologues were used as a template. Typical RING E3 ubiquitin ligase structures include: a distinct cross-brace (or RING finger) motif (Freemont *et al.*, 1991) with two zinc ion centres (connected by two loops) and a central alpha helix, forming an interface for interactions with E2 ubiquitin conjugating enzymes (Chapter 1, Figure 1.5) (Metzger *et al.*, 2014).

Both RING models (10 and 106) include a single, central alpha helix (A326-R335); preceded by two antiparallel beta strands (β 1, C314- C317; β 2, M321- C325) (Figure 3.20, Panel A). In addition, RING model 10 includes a third beta strand (G342-N349), with an additional loop region (up to residue L363); predicted by iterative modelling. Whilst a consensus between PBD template files was reached, regarding the presence of both antiparallel beta strands, JPred (Cuff and Barton, 2000) and PSI-Pred (Buchan *et al.*, 2013) did not predict either secondary structural features (Figure 3.4). This suggest that secondary structural predictions by JPred and PSI-Pred, derived from sequence and structural evolutionary information respectively, may be slightly less accurate in comparison to extrapolation of known secondary structural information from experimentally derived template (and therefore homologous) structures.

3.3.2 Loop conformation and interface residue orientation may be sources of error within the CHFR RING models

Ab initio loop refinement of RING model 106 (spanning alignment position 6 to 16) was required, since an increase in DOPE energy at the protein N-terminus was observed, in comparison to the best multi-template model obtained (Figure 3.6). Interestingly, variation in loop conformation spanning this region corroborates with the significant changes in RMSF values. This may be indicative of lower protein stability within this region, or loop refinement errors, since regions of lowest residue conservation that lack secondary structural content (as observed within loop regions) are more likely to be regions of lower model reliability. Loop regions at the CHFR RING model N-terminus, C-terminus and between zinc binding motifs exhibit the lowest percentage of sequence similarity. This is indicative of less reliable regions of the CHFR RING model, attributed to a lack of sequence and structural similarity between the target and template sequences.

Homology modelling of protein structures is severely limited in the presence of variable, non-homologous regions where no suitable templates can be used, most particularly in loop regions (Moult *et al.*, 2005; 2013; Kinsch and Grishin, 2002). By measuring local superimposition of homology modelling-produced and experimental structures, the average accuracy of predicting the main chain decreases with increasing loop size (Fiser *et al.*, 2000). For example; 4, 8 and 12 residue loops can be predicted in MODELLER with an accuracy of 100, 90 and 30%, respectively (Fiser *et al.*, 2000). Taking this into consideration, it is unsurprising that the CHFR RING models 106 (alignment position 6 to 16, N-terminus) and 10 (alignment position 58 to 82, C-terminus) exhibited higher DOPE profiles, greater variability in loop-refined models produced and RMSF fluctuations within these areas. Extrapolating the findings of Fiser *et al.* (2000), the extended RING model 10 C-terminal loop of 10 residues may correspond to an accuracy of ~24%, (or 1.91 ± 0.13 Å average accuracy). Assessment of multiple loop building computational tools revealed MODELLER was marginally better, in comparison to

SWISS-model, Builder, Nest and SegMod/ENCAD (Dalton and Jackson, 2007), suggesting the loop regions produced within the CHFR RING models are best representative of possible structures.

3.3.3 Modelling the CHFR RING stacking interaction

Whilst RNF4 forms a dimer via the third alpha helix (Plechanovova *et al.*, 2012; Figure 3.15), this secondary structural feature is absent from both CHFR RING models. Supporting the hypothesis that CHFR multimer formation (most likely, as a dimer) is required for optimum E3 ubiquitin ligase activity, the second CHFR RING model 10 (produced via iterative modelling) exhibited a high sequence identity with RNF4 (30.6%, SIAS server) and good stereochemical quality. Through the re-evaluation of the CHFR: RNF4 interaction in light of the iterative modelling step used to produce RING model 10 (Figure 3.22), a stacking interaction between ubiquitin (G35) and two different CHFR RING domains (contributing either H322 or Y362 within the interaction) was modelled. With CHFR: Ubc13 interactions modelled using RING model 10 identical to those predicted in RING model 106, multimer formation mediating the CHFR interaction with Ubc13 *and* UbcH5a may be also be required for CHFR E3 ubiquitin ligase activities.

3.3.4 Molecular dynamic (MD) simulations indicate good fold and stability of the RING model backbone, with some N- and C-terminal flexibility

In an attempt to predict the structural stability of both models, MD simulations were also implemented. The CHFR RING model 10 radius of gyration (R_g) and RMSD trajectories appeared to be stabled within the last 50ns of MD simulations (Figure 3.25); with a slightly higher density identified (1016.32Kg m^{-3}) in comparison to the TIP3P water model density at 300K (Mahoney and Jorgensen, 2000). Since the CHFR RING model exhibits good stereochemical quality and validity comparable to PDB structures, a lack of stability within the CHFR RING domain C-terminus is

inevitable, considering the lack of secondary structural content (with the exception of the third beta strand) predicted within this region.

The ZAFF AMBER force field (Peters *et al.*, 2010) utilizes the Van der Waals parameters for zinc ions developed by Li *et al.* (2013). Within the non-bonded model, reproducibility of +2 metal cation (M(II)) properties in aqueous solutions is difficult; since interactions between the metal ion and surround water can be underestimated (Li *et al.*, 2013). Considering that the experimental properties of M(II) metal ions are water model dependent (Li *et al.*, 2013) reproduction of the MD experiments described here using additional water models may further support the conclusions drawn regarding predicted CHFR RING model stability. In addition, removal of zinc metal ion centres from the original PDB and generated topology files could be used to explore the behaviour (RMSD, RMSF) and fold (R_g) of the CHFR RING domain in solution and the specific role of the metal ion centres.

**CHAPTER 4: Verification of interactions within
the CHFR RING: Ubc13 (: Mms2) ~ ubiquitin
complex**

INTRODUCTION

Numerous domain deletion mutants of the CHFR central RING domain used within ubiquitination assays have verified it as essential in mediating its *in vivo* and *in vitro* ubiquitination activities (Bothos *et al.*, 2003; Scolnick and Halazonetis, 2000; Chaturvedi *et al.*, 2002). Previous experiments have sought to identify evolutionary conserved RING domain residues between CHFR and its homologues (Kang *et al.*, 2002; Scolnick and Halazonetis, 2000), with site directed mutants exhibiting ablation of RING E3 activities in mediating polyubiquitin chain formation. However, no structural insight into these specific interactions has been established in verifying the entire CHFR RING: E2 and CHFR RING: ubiquitin interface residues and interactions responsible for these protein: protein interactions.

By exploring experimental X-ray structures of RING E3 homologues in complex with E2~ ubiquitin and E2: pseudo E2 enzymes, the interactions within the CHFR RING: Ubc13 ~ ubiquitin and CHFR RING: Ubc13 (: Mms2) protein complexes have been modelled. In addition, site directed FL-CHFR (RING) and ubiquitin mutants have been used to provide, for the first time, an in-depth analysis and verification of key CHFR RING residues essential in mediating polyubiquitin chain formation *in vitro*.

4.1 MATERIALS AND METHODS

4.1.1 Modelling interactions between the CHFR RING domain, Ubc13 E2 and ubiquitin

To deduce CHFR RING domain residues most likely to interact with either the Ubc13 E2 ubiquitin conjugating enzyme or ubiquitin within the CHFR RING: Ubc13 (: Mms2) ~ Ub protein complex, the CHFR RING model (10) cartoon was superimposed with RING E3 ubiquitin ligases within experimental (PDB-derived) E3: E2 ~ ubiquitin crystal structures in Pymol (Version 1.8.0.5; Schrödinger, 2015).

Interactions were then cross-referenced with the original multiple sequence alignments between the CHFR RING domain and PDB local database structures used to produce the original CHFR RING homology model.

Selected structures containing a RING E3 ubiquitin ligase, ubiquitin, Ubc13 and (occasionally) Mms2 from the original PDB BLAST search were then assessed, in more detail, for sequence and structural similarity with the CHFR RING domain. In addition, an annotated multiple sequence alignment (using JalView and Clustal Omega, with manual adjustment; Clamp *et al.*, 2002 and Sievers *et al.*, 2011) between E2s and the CHFR RING domain informed of secondary structure content within the RING model and respective E3 homologues (Chapter 3, Figure 3.4).

PISA (Version 1.51; Krissinel and Henrick, 2007) was then used to model (in more detail) potential protein-protein interaction types and distances between the CHFR RING model (10) with Ubc13 and the CHFR RING model with ubiquitin (Appendix 3, Tables A3.1 to A3.6, inclusive).

A selected number of modelled interactions between the CHFR RING domain or other E3 ubiquitin ligases with Ubc13 and ubiquitin were well conserved, and therefore deduced as important in producing the Ubc13: CHFR RING and ubiquitin: CHFR RING interfaces.

Therefore, interactions between the CHFR RING domain side chain (non-buried) residues and either Ubc13 or ubiquitin were then targeted for site directed mutagenesis, in an attempt to disrupt such interactions. This ablated the formation of the CHFR RING: E2 ~ ubiquitin interface and were used to identify its effect on polyubiquitin chain formation via ubiquitination assay and western blot analysis.

4.1.2 Site directed mutagenesis of full-length CHFR

4.1.2.1 Two-step PCR

Selected interactions modelled as important at either the CHFR RING: Ubc13 or CHFR RING: ubiquitin interfaces and most likely to be essential in E3: E2: ubiquitin complex formation, prior to polyubiquitin chain formation, were identified (Section 4.1.1).

In order to disrupt such interactions and therefore verify their importance in complex formation and CHFR-specific polyubiquitin chain formation, site directed mutagenesis was performed, mutating surface (and therefore non-buried) CHFR RING residues to disrupt interactions between the CHFR RING domain and Ubc13 or ubiquitin.

PCR primers were designed to introduce specific mutations within codons specific to targeted CHFR RING residues (Appendix 3, Table A3.7). All mutations introduced an alanine residue at selected positions within the CHFR RING domain, with the exception of H322. Here, a cysteine was introduced to retain coordination of at least 4 residues with the second zinc metal ion binding centre whilst disrupting the Y363-H322 (CHFR RING): G35 (ubiquitin) stacking interaction previously modelled.

Primers were designed to target a codon and mutations included in each individual primer per primer pair. Primers had a total length of 25 to 45 bases, an annealing/melting temperature (T_m) no lower than 78 °C; with least 40 % GC content (Wang and Malcolm, 1999). An equal number of bases (10 to 15) flanked the targeted codon, with primers terminated with one or more G or C bases (Wang and Malcolm, 1999). The minimum number of new bases was introduced within the targeted codon to encode for the new amino acid.

To avoid primer dimer formation, a two-step PCR reaction protocol was adapted (Wang and Malcolm, 1999). Reactions were set up (in duplicate) per primer pair (Appendix 3, Table A3.7) within 0.5 mL thin-walled PCR tubes; consisting of 1 pmol of forward or reverse primer (Integrated DNA Technologies UK Limited; desalted

and unmodified), 50 ng template (wild-type full-length Strep (II)-CHFR), 2.5 μ L (5 % v/v) DMSO, 0.2 mM each dNTP (premixed, 10 mM stock, NEB), 5 μ L *Pfu* Turbo polymerase reaction buffer (10 X stock, Agilent) and deionised water to a total volume of 49 μ L. 2.5 units of *Pfu* Turbo DNA polymerase (1 μ L) were then added and tubes gently mixed before a brief centrifugation (11,000 g, 30 seconds, 4 °C).

The first stage of the PCR reaction was then carried out within a thermocycler (Techine TC-3000, Bibby Scientific) as described in Table 4.1, with the lid preheated throughout (105 °C) and reactions held at 4 °C before proceeding with the second stage.

Table 4.1 Thermocycler reaction conditions for the first stage of PCR-mediated site directed mutagenesis of full-length CHFR double-stranded DNA (1-664).

Segment	Description	Temperature (°C)	Time (minutes : seconds)	Cycles
1	Initial denaturation	95.0	1:00	1
2	Denaturation	95.0	0:30	10
	Annealing	55.0	1:30	
	Elongation	68.0	15:00	
3	Hold	4.0	∞	-

Following completion of the first stage, each reaction pair (corresponding to their respective targeted codon and therefore mutation) were combined, with each 50 μ L reaction pair added to a single 0.5 mL thin-walled PCR tube. 0.5 μ L *Pfu* turbo (1.25 units) was then added, and tubes gently mixed before a brief centrifugation (11,000 g, 30 seconds, 4 °C).

Each site directed PCR reaction was then subjected to a second stage of PCR reactions within the thermocycler (Table 4.2); with the lid heated throughout (to 105 °C) and reactions held at 4 °C prior to *DpnI* digestion.

Table 4.2 Thermocycler reaction conditions for the second stage of PCR-mediated site directed mutagenesis of full-length CHFR double-stranded DNA (1-664).

Segment	Description	Temperature (°C)	Time (minutes : seconds)	Cycles
1	Initial denaturation	95.0	1:00	1
2	Denaturation	95.0	0:30	18
	Annealing	55.0	1:30	
	Elongation	68.0	15:00	
	Hold	4.0	∞	

Following the completion of the second PCR reaction, 2 μ L (2 units) of *DpnI* were added and tubes gently mixed before a brief centrifugation (11,000 g, 30 seconds, 4 °C). Reactions were then incubated at 37 °C for 1 hour, and then incubated on ice for 2 minutes prior to transformation of *E. coli* DH5 α for DNA expression.

4.1.2.2 Transformation of *E. coli* DH5 α competent cells

DH5 α cells were then transformed by initially thawing 20 μ L of competent cells on ice; followed by addition of 1 μ L *DpnI* –treated double-stranded PCR product. Cells were then incubated on ice for 30 minutes, followed by a brief heat-shock at 42 °C for 42 seconds. Transformed cells were then incubated on ice for 2 minutes, proceeded by the addition of 200 μ L LB and incubation at 37 °C for 1 hour (37 °C, 5.844 x g; Infors HT, Ecotron). Following the transformations, 200 μ L of each transformation was spread across LB-agar plates (containing 50 μ g/ mL kanamycin only) and incubated overnight at 37 °C.

The following day, 5 mL LB (with Kanamycin at a 50 μ g/ mL working concentration) was inoculated with a single colony and incubated overnight (37 °C, 5.844 x g; Infors HT, Ecotron).

4.1.2.3 Purification of plasmid DNA

5 mL of inoculated LB was briefly centrifuged (5 minutes at 4°C) within a SX4750 swinging bucket rotor at 2,890 x *g* (3,750 RPM at r_{\max} of 207.8 mm; Allegra X-12R centrifuge, Beckman Coulter); with the supernatant discarded.

A QIAprep Spin Mini Prep Kit (Qiagen) was then used to purify the plasmid DNA, according to manufacturer's instructions and eluted with a single 50 μ L volume of ultrapure water.

4.1.2.4 Quantification of purified DNA

DNA concentrations were determined by pipetting 1.5 μ L eluted DNA within a Nanodrop-2000 Spectrophotometer (ThermoScientific) and measuring the absorbance at 260 nm.

All purified DNA was stored at – 20 °C prior to use.

4.1.2.5 DNA sequencing

Constructs produced using the two-step PCR site directed mutagenesis protocol were then subjected to sequencing, verifying the presence of mutations introduced within the specific CHFR RING domain codons (Appendix 3, Table A3.7).

A sequencing primer, complimentary to a region just outside the full-length CHFR protein's central RING domain (5' →3', CGTCCACGAGGACGTCAG), was designed in accordance to recommendations for ValueRead Sequencing Services (Eurofins).

ValueRead sequencing (Eurofins) was performed using individual DNA constructs and the CHFR sequencing primer to confirm the presence of the mutation via a

single sequencing reaction per mutation; with a 900 to 1000 base coverage typically achieved.

Sequenced constructs (.txt files) were then compared to the counterpart wild type full-length CHFR sequence by aligning both sequencing within SerialCloner (Version 2.6.1); with the mutation of selected codons verified prior to proceeding protein expression and purification.

4.1.2.6 Expression and purification of FL-CHFR RING mutant proteins

Full-length CHFR RING mutant proteins were expressed and purified as previously described (Chapter 2, Section 2.1.2); with proteins resolved by SDS-PAGE, as previously described (Chapter 2, Section 2.1.2.3).

Identical SEC conditions were used to express and purify the original FL-CHFR protein, domain deletion mutant proteins and RING domain site directed mutants. An identical procedure for calculation of molecular weights of RING domain mutant proteins and analysis of chromatograms was therefore also used to determine whether any RING domain specific mutations affected the oligomeric state of the FL-CHFR mutants (Chapter 2, Section 2.1.3.1).

All proteins were flash frozen in liquid nitrogen and stored at – 80 °C.

4.1.3 Site directed mutagenesis of ubiquitin

4.1.3.1 Two-step PCR

PCR reactions were performed as previously described (Section 4.1.2.1), using primers (Integrated DNA Technologies UK Limited; desalted and unmodified) specifically complimentary and targeting ubiquitin residues (Appendix 3, Table A3.8). The mutations were targeted to specifically disrupt CHFR RING – ubiquitin

interaction and ablate polyubiquitin chain formation. Different thermocycler conditions for the first (Table 4.3) and second (Table 4.4) stage PCR reactions, accounting for differences in plasmid size between the full-length CHFR and ubiquitin-encoding constructs, are indicated below.

Table 4.3 Thermocycler reaction conditions for the first stage of PCR-mediated site directed mutagenesis of ubiquitin.

Segment	Description	Temperature (°C)	Time (minutes : seconds)	Cycles
1	Initial denaturation	95.0	1:00	1
	Denaturation	95.0	0:30	
2	Annealing	55.0	1:30	10
	Elongation	68.0	8:00	
3	Hold	4.0	∞	-

Table 4.4 Thermocycler reaction conditions for the second stage of PCR-mediated site directed mutagenesis of ubiquitin.

Segment	Description	Temperature (°C)	Time (minutes : seconds)	Cycles
1	Initial denaturation	95.0	1:00	1
	Denaturation	95.0	0:30	
2	Annealing	55.0	1:30	18
	Elongation	68.0	8:00	
3	Hold	4.0	∞	-

4.1.3.2 *E. coli* DH5α transformation, DNA purification and sequencing

Chemically competent DH5α *E. coli* was transformed with the *DpnI*-incubated DNA (Section 4.1.2.2); with DNA purified (Section 4.1.2.3) and quantified (Section 4.1.2.4), as previously described.

A (universal) sequencing primer (Eurofins), complimentary to the N-terminus of the pET-17 vector (5' → 3', GCA TCA CCA TCA CCA TC), was then used to sequence the constructs, as previously described (Section 4.1.2.5).

4.1.3.3 Expression and purification of His6-ubiquitin

Wild type and mutant His6-ubiquitin proteins were expressed and purified by metal ion affinity chromatography via a HisTrap FF column (via manual syringing; GE Healthcare); as previously described for all E2 ubiquitin-conjugating enzymes (UbcH5a, Ubc13 and Mms2; Chapter 2, Section 2.1.6.1); with proteins resolved by SDS-PAGE, as previously described (Chapter 2, Section 2.1.2.3; Appendix 3, Figure A3.9). Proceeding washing of unbound proteins, mutant Ubc13 E2 proteins were eluted using a single-step 15 mL 250 mM imidazole application.

All proteins (in 5 % v/ v glycerol) were flash frozen in liquid nitrogen and stored at – 80 °C.

4.1.4 Ubiquitination assays using FL-CHFR and ubiquitin site directed mutants

Ubiquitination assays using FL-CHFR and ubiquitin mutant proteins (targeting interactions between the CHFR RING domain and Ubc13 E2 enzyme or CHFR RING domain and ubiquitin) were carried out as previously described, with counterpart WT (FL-CHFR and ubiquitin) proteins also included.

Assay reactions were then resolved by SDS-PAGE (Chapter 2, Section 2.1.2.3) and polyubiquitin chain formation detected by performing western blots against ubiquitin (Chapter 2, Section 2.1.2.7).

4.2 RESULTS

4.2.1 interfaces within the CHFR RING: Ubc13 ~ ubiquitin complex

To model residues at the CHFR: E2 heterodimer (Ubc13: Mms2) interface, the final CHFR RING model (10) cartoon was aligned (in PyMol) with the deposited heretrotrimeric complex of RNF8: Ubc13 (:Mms2) (PDB: 4ORH, Campbell *et al.*,

2012) (Figure 4.1, Panel B). In addition, a multiple sequence alignment and secondary structural annotations were cross-referenced with the aligned cartoons (Figure 4.1, Panel A).

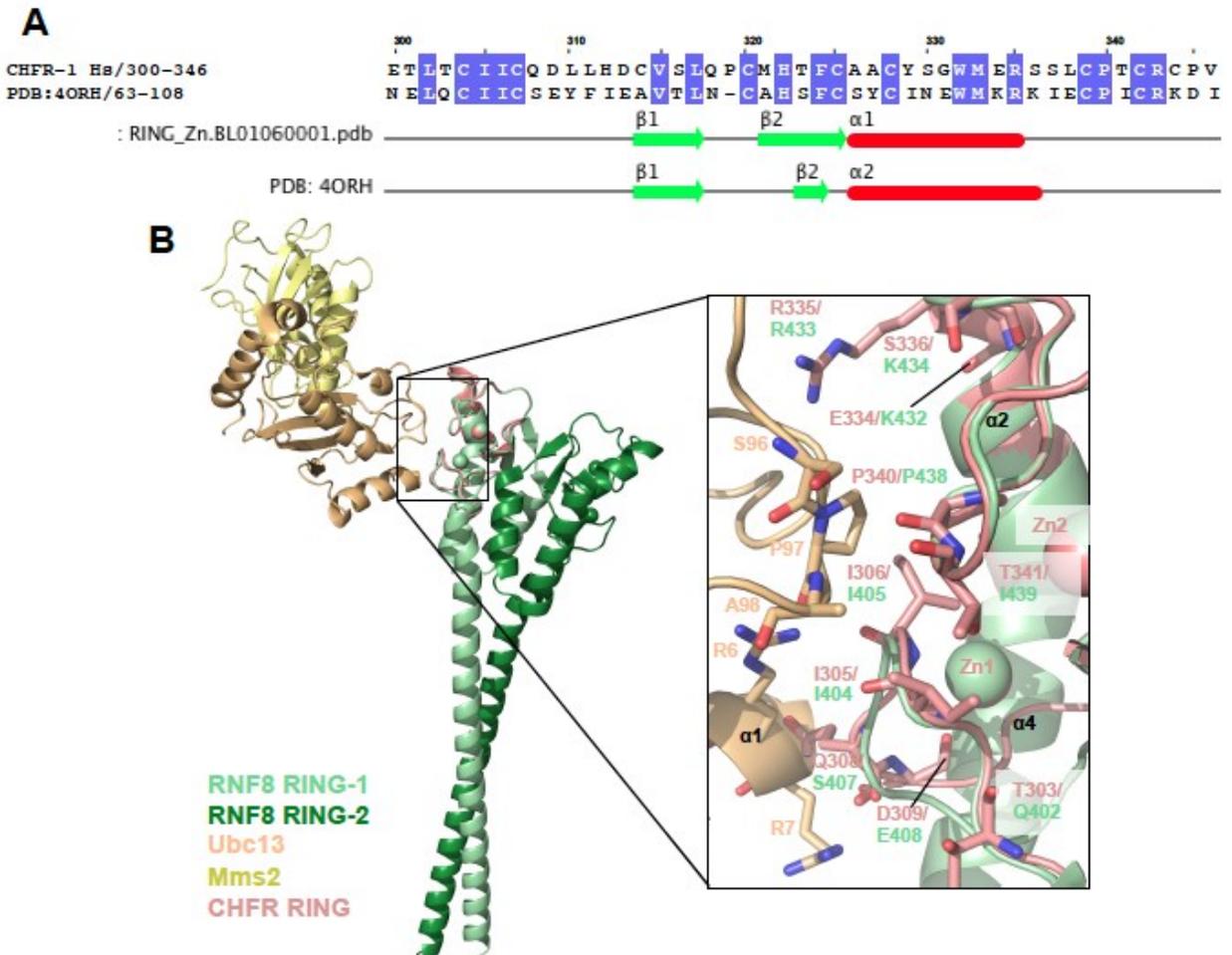


Figure 4.1 Alignment between the final CHFR RING model (10) and the E3 ubiquitin ligase homologue RNF8 (PDB: 4ORH, Campbell *et al.*, 2012). (A) A multiple sequence alignment between RNF8 and CHFR RING domain corresponds to a 41.03% sequence identity, with identical residues coloured in purple and amino acid numbering used corresponding to the CHFR RING- domain (Isoform 1). Both structures share a significant amount of secondary structure content, including two antiparallel beta strands (C304 to C307 and M326 to C325, CHFR RING, green arrows) and a central alpha helix (A326 to R335, CHFR RING, red line). (B) The cartoon of the CHFR RING model was then aligned with RNF8 and key residues at the CHFR: Ubc13 interface extrapolated from the original RNF8: Ubc13 (:Mms2) crystal structure and shown as sticks (Campbell *et al.*, 2012). Corroborating with the findings regarding RNF8: Ubc13 interactions by Campbell *et al* (2012), CHFR RING hydrophobic residues P340, I305 and I306 may be in a position to make electrostatic contacts with the E2- conserved SPA motif (96, 97 and 98, respectively) of Ubc13. T341 may not to participate in this interaction, in comparison to observations in RNF8 (I349) (Campbell *et al.*, 2012). Zinc ions are shown as spheres.

Campbell *et al* (2012) verified (by mass spectrometry) that RNF8 interacts exclusively with Ubc13 within the tertiary complex and not Mms2, corroborating with the interaction specificity modelled between the CHFR RING domain and Ubc13. Despite the low resolution of the crystallographic data set (PDB: 4ORH, 4.8Å), an interface between the Ubc13 (E2 conserved) SPA motif (96, 97 and 98, respectively) and RNF8 (Campbell *et al.*, 2012) can be extrapolated to conserved residues present in the CHFR RING domain (P340, I305 and I306). However, the dimer interface present in RNF8 (corresponding to C-terminal α 3) (Campbell *et al.*, 2012) is absent from the CHFR RING domain.

In addition, in-depth analysis of the RNF8: Ubc13 interfaces within the X-ray experimental structure (Campbell *et al.*, 2012; PDB: 4ORH) using PISA; cross-referenced and compared to the respective homologous residues within the CHFR RING model, provided more detail regarding the modelled CHFR RING: Ubc13 interface (Appendix 3, Tables A3.1 to A3.3, inclusive). Within the CHFR RING: Ubc13 model, hydrogen bonds between CHFR RING residues I305 and I306 with R7 of Ubc13 are homologous to RNF8 RING residues I404 and I405, respectively (Figure 4.1, Panel B). In addition, Q308 and D309 may form hydrogen bonds with Ubc13 residues R6 and R7 (Figure 4.1, Panel B). M64 of Ubc13 was also modelled to form a hydrogen bond with CHFR RING R335, homologous to the experimentally observed RNF8 RING residue R433 (Appendix 3, Tables A3.1 to A3.3, inclusive; Figure 4.1, Panel B).

Superimposition of CHFR RING model 10 with RNF4 within the RNF4 (fused dimer): Ubc13 ~ polyubiquitin C (crystal) complex experimentally obtained by Branigan *et al* (2015) provided additional insight into predicted interactions at the CHFR RING: ubiquitin and CHFR RING: Ubc13 interfaces (Figure 4.2; Appendix 3, Tables A3.4 to A3.6, inclusive).

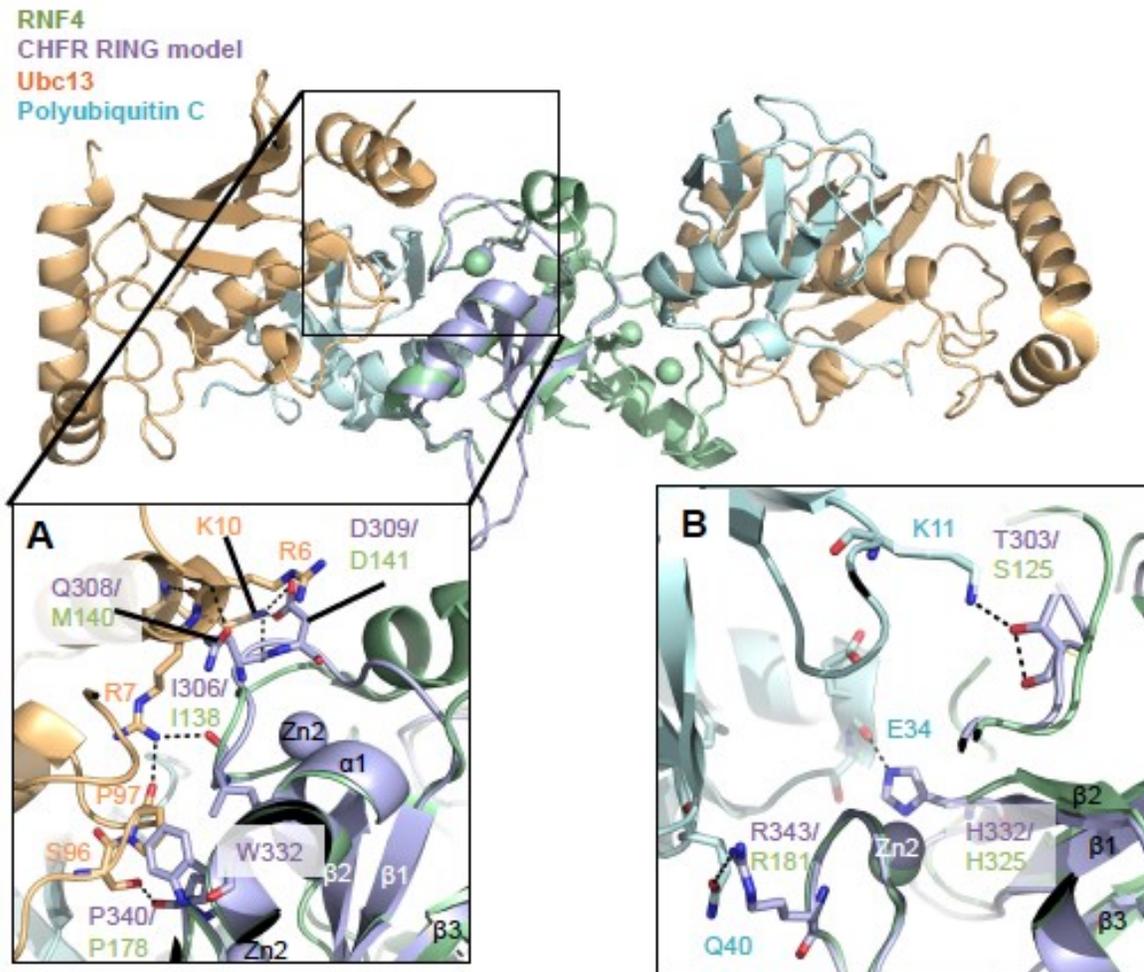


Figure 4.2 Structure of the CHFR RING model superimposed upon the E3 homologue RNF8, in complex with polyubiquitin C and Ubc13 (using experimental dataset; PDB: 5AIU, Branigan *et al.*, 2015). The CHFR RING model was superimposed upon the RNF 4 fused dimer, within the experimental dataset from Branigan *et al* (2015). (A) Detailed (modelled) interface between the CHFR RING domain homology model (pale purple) and Ubc13 (orange); with respective RNF8 homologous residues (green) also indicated. Hydrogen bond and salt bridge formation was modelled between CHFR RING Q308 (homologous to RNF4 M140) and Ubc13 residues R6 and K10. In addition, hydrogen bond formation between the CHFR RING P340 (homologous to RNF4 P178) and Ubc13 residue S96 was also modelled. As modelled via the CHFR RING: Ubc13 PDB 4ORH (Figure 21), electrostatic interactions between P340 of the CHFR RING (or P178 of RNF4) were also modelled. (B) Molecular detail of the polyubiquitin C (pale cyan) and CHFR RING (pale blue) interfaces (with identical box positioning to section (A); with RNF8 homologous residues also indicated (green) and interface positioned behind the Ubc13 protein (removed for clarity). Hydrogen bond formation between CHFR RING H322 (or equivalent RNF4 residue H325) and ubiquitin E34 was also modelled. In addition, hydrogen bond formation between ubiquitin residue Q40 and CHFR RING R343 (or R181 in RNF4) was also modelled. Ubiquitin residue K11 was also predicted to interact with T303 of the CHFR RING domain.

A stacking interaction was also modelled between residues H322 of one CHFR RING domain and Y362 of a second; with G35 of ubiquitin also participating in the interaction (Figure 4.3); strongly comparable to the stacking interaction predicted using the RNF4: Ubch5a ~ ubiquitin experimental X-ray structure (PDB: 4AP4, Plechanovová *et al.*, 2012).

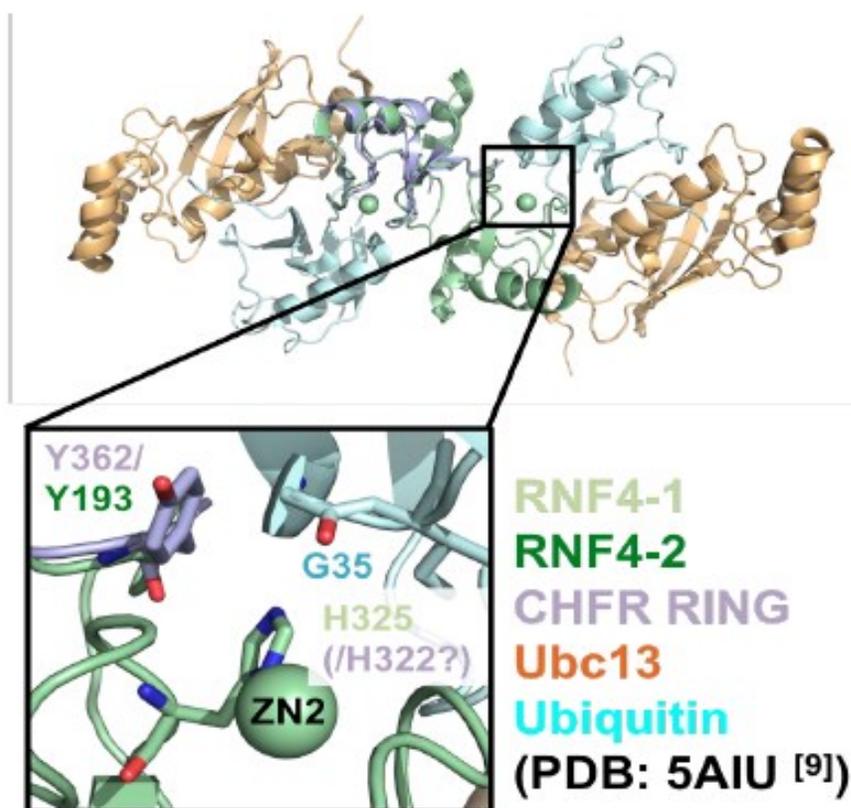


Figure 4.3 Stacking interaction modelled between two individual CHFR RING domains and ubiquitin. The CHFR RING model was superimposed onto the structure of the E3 ubiquitin ligase homologue RNF4, within the experimental RNF4 (fused dimer): Ubc13: polyubiquitin C (or *ubiquitin*) complex (Branigan *et al.*, 2015; PDB: 5AIU). The Y362 residue of one CHFR RING domain (CHFR RING-1) may stack onto that of H322, derived from a second CHFR RING domain (CHFR RING-2); with G35 of ubiquitin also at the interface.

In-depth analysis of the RNF4: Ubc13 and RNF4: polyubiquitin C interfaces within the X-ray experimental structure (Branigan *et al.*, 2015; PDB: 5AIU) using PISA; cross-referenced and compared to the respective homologous residues within the CHFR RING model, provided more detail regarding the modelled CHFR RING: Ubc13 interface (Appendix 3, Tables A3.4 to A3.6, inclusive).

Modelling the CHFR RING: Ubc13 and CHFR RING: ubiquitin interfaces was therefore achieved by assessing two independent X-ray structures containing E3 ubiquitin ligase homologs. These predicted protein-protein interactions would form the basis of *in vitro* ubiquitination experiments to verify residues essential for CHFR- specific polyubiquitin chain formation.

4.2.2 The majority of FL-CHFR RING domain mutant proteins are dimeric in solution, with MW comparable to the FL-CHFR WT counterpart

Interactions at the CHFR RING domain extended model (10) and Ubc13 interface, modelled via two independent experimental structures of E3: E2 (: pseudo E2) and E3: E2 ~ ubiquitin protein complexes (Section 4.2.1), were then compiled.

Surface (non-buried) residues were identified within the CHFR RING domain, with site directed mutants expressed, purified and used within ubiquitination assays to target key protein-protein interactions at the RING: Ubc13 interface. Mutations included: R345A, R343A, R335A, D309A, I306A, T303A, L363A, W332A, Q308A, P340A, H322C/ Y362A, I305A, E300A, Y362A and H322C.

Identical purification strategies were used for RING domain site directed mutants in comparison to all other purified FL-CHFR proteins. Therefore, an identical molecular weight standard standard curve and respective calculations were used compare molecular weights of RING domain site directed full-length CHFR mutants, in comparison to wild-type counterparts (Figure 4.4, Table 4.5).

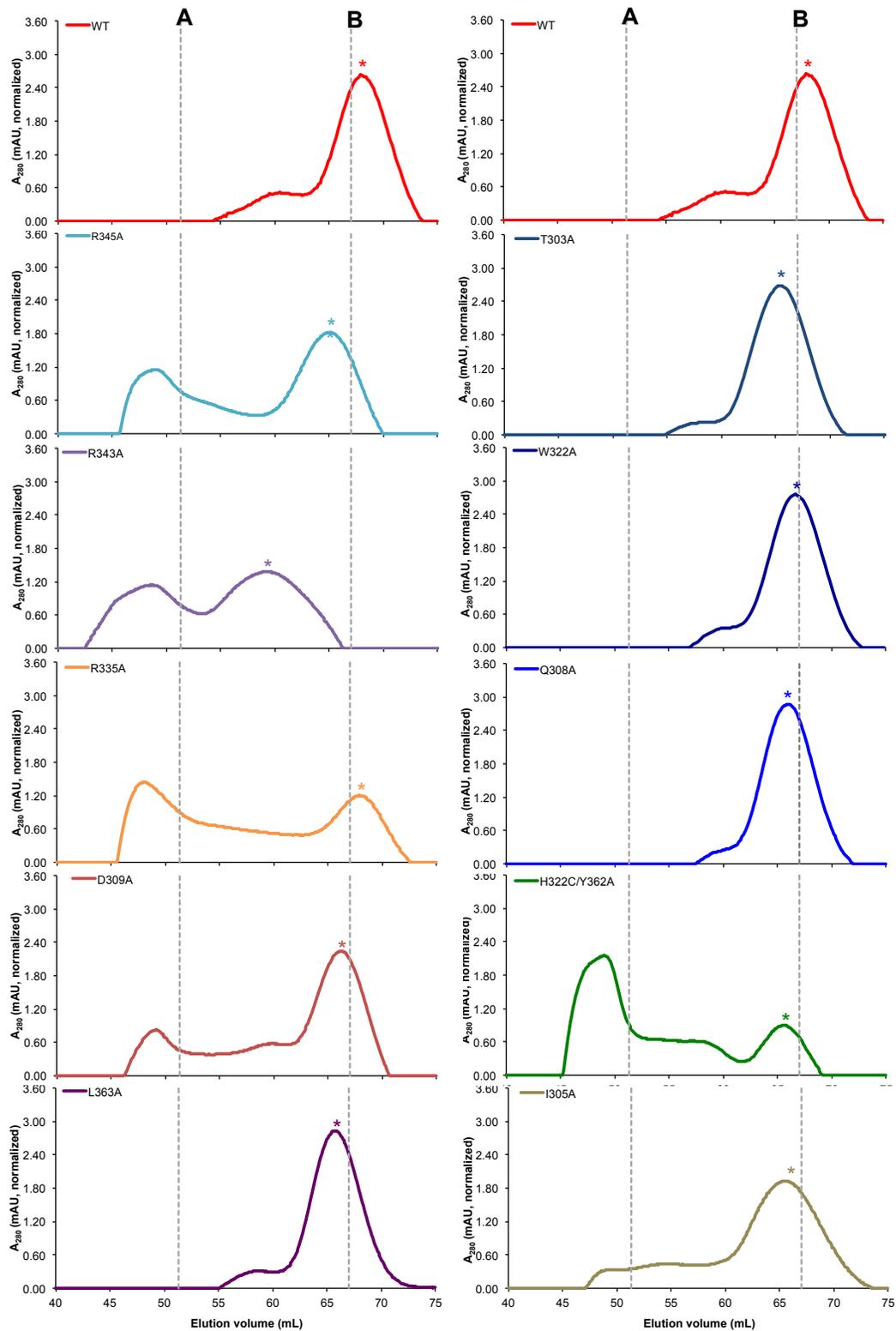


Figure 4.4 Chromatograms of SEC purified FL-CHFR RING site directed mutants with molecular weight standards (Superdex 200 16/600; GE Healthcare). Identical chromatograms, corresponding to the wild-type FL-CHFR protein, are shown at the top (for comparative purposes) in red. Elution volumes of specific molecular weight standards are indicated by the vertical grey lines, corresponding to A: Thyroglobulin (bovine, 670 kDa) and B: γ -globulin (bovine, 158 kDa).

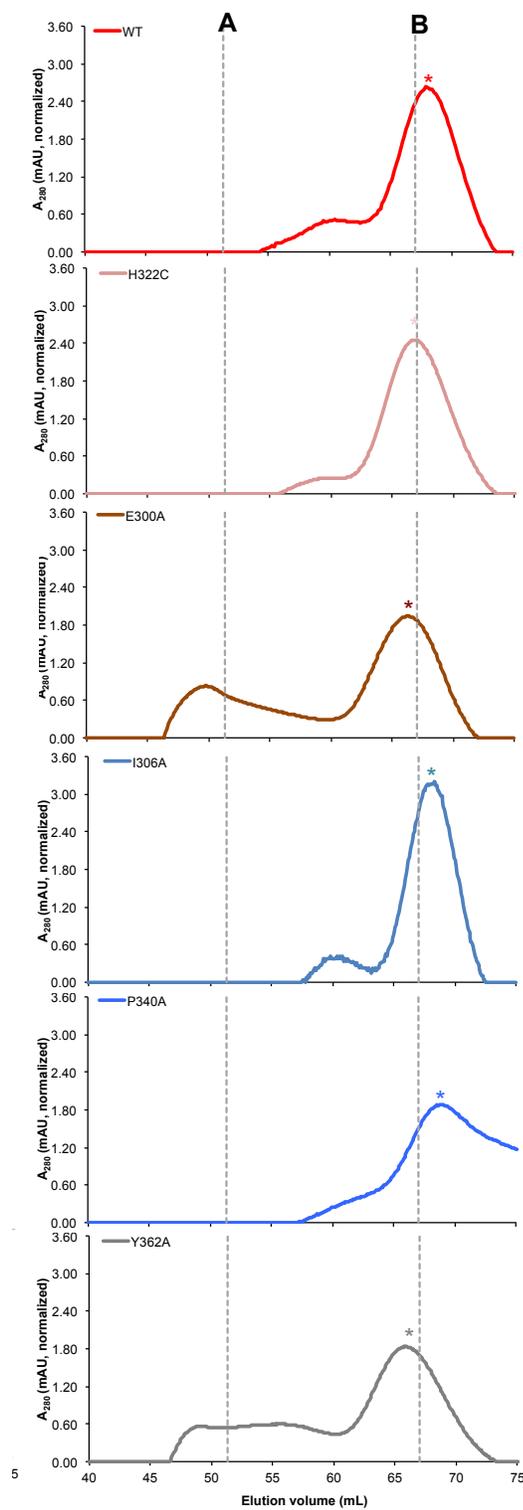


Figure 4.4 (Continued...) Chromatograms of SEC purified FL-CHFR RING site directed mutants with molecular weight standards (Superdex 200 16/600; GE Healthcare). Identical chromatograms, corresponding to the wild-type FL-CHFR protein, is shown at the top (for comparative purposes) in red. Elution volumes of specific molecular weight standards are indicated by the vertical grey lines, corresponding to A: Thyroglobulin (bovine, 670 kDa) and B: γ -globulin (bovine, 158 kDa).

Table 4.5 Calculated molecular weights of CHFR RING domain mutant proteins. Molecular weights were calculated using SEC molecular weight standards (Biorad) via a Superdex S200 16/600 SEC column (GE Healthcare) and calibration curve, as previously described. Shaded boxes indicate lowest Δm % values, and therefore correspond to the correct oligomeric state.

	Elution volume	MW (kDa)	Δm %: dimeric (158.83 kDa)	Δm %: quaternary (317.64 kDa)	Predicted oligomeric state
WT	67.91	154.72	- 2.57	+ 51.04	Dimeric
R345A	65.09	205.62	+ 29.48	+ 34.93	Dimeric
R343A	59.12	375.44	+ 136.43	+ 18.81	Quaternary
R335A	67.91	154.72	- 2.57	+ 51.04	Dimeric
D309A	66.26	182.73	+ 15.07	+ 42.17	Dimeric
I306A	68.29	148.90	- 6.23	+ 52.88	Dimeric
T303A	65.53	196.69	+ 23.86	+ 37.76	Dimeric
L363A	65.71	193.15	+ 21.83	+ 38.88	Dimeric
W332A	66.84	172.34	+ 8.53	+ 45.46	Dimeric
Q308A	65.68	193.74	+ 22.00	+ 38.69	Dimeric
P340A	68.73	142.44	- 10.30	+ 54.92	Dimeric
H322C/ Y362A	65.68	193.74	+ 22.00	+ 38.69	Dimeric
I305A	65.50	197.29	+ 24.24	+ 37.57	Dimeric
E300A	66.37	180.71	+ 13.80	+ 42.81	Dimeric
Y362A	65.33	200.70	+ 26.38	+ 36.49	Dimeric
H322C	66.71	174.62	+ 9.69	+ 44.74	Dimeric

MW = molecular weight; *M* = monomer/ monomeric; *D* = dimer/ dimeric.

[1] Predicted molecular weights calculated using protein amino acid sequence within the ProtParam online tool (Gasteiger *et al.*, 2002).

[2] Molecular weights calculated using calibration curve produced using size exclusion chromatography and molecular weight standards; with known elution volumes of proteins (*x*) fitted to the curve $y = -0.0438x + 8.164$; where *y* is the calculated molecular weight.

[3] [Δm %] indicates percentage differences with respect to the calculated (theoretical) mass (see [1]).

[4] Deduced oligomeric and quaternary states correspond to calculated MW with lowest percentage difference [Δm %] relative to predicted MW (see [1]).

The majority of molecular weights calculated for FL-CHFR (RING) site directed mutants correspond to the wild-type counterpart, indicative of their dimeric oligomeric state upon elution via SEC. However, the elution volume and corresponding calculated molecular weight of the FL-CHFR (RING) mutant protein R343A is closer to a quaternary oligomerization state, rather than the expected dimeric counterpart. Since the percentage difference in theoretical MW of the

quaternary state in comparison to calculated molecular weight is considerably lower, in comparison to dimeric counterpart (Table 4.5, +18.81 compared to +136.43%), the R343A protein is most likely to adopt a quaternary state in solution, at the (specific) SEC eluted concentration and conditions.

4.2.3 Verification of modelled interactions at the CHFR RING: Ubc13 and CHFR RING: ubiquitin interfaces

To verify CHFR RING domain residues essential in forming the E3: Ubc13 and E3: ubiquitin interfaces required for polyubiquitin chain formations, interactions were modelled using PDB deposited experimental structures, combined with PISA bioinformatics analysis (Section 4.1.1). FL-CHFR RING site directed mutant proteins were then recombinantly expressed and purified, with the majority (except R335A) exhibiting a similar elution and molecular weights corresponding to a dimeric E3 ubiquitin ligase, comparable to the WT FL-CHFR counterpart.

Ubiquitination assays, including the FL-CHFR RING mutant proteins (and FL-CHFR WT counterpart controls), WT ubiquitin, WT Ubc13 and WT Mms2, were then used to assess the effects of targeting key interface residues *in vitro* upon polyubiquitin chain formation (Figure 4.5).

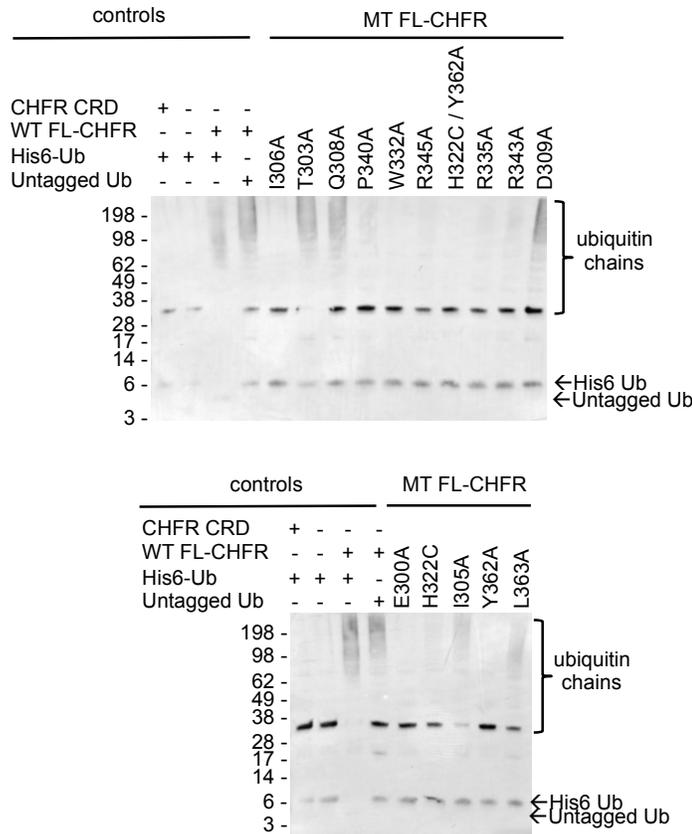


Figure 4.5 Mutational analysis of the FL-CHFR: Ubc13 (:Mms2)~ Ub complex using CHFR RING site directed mutants. Ubiquitination assays were performed using site-directed FL-CHFR (RING) mutant proteins with WT Ubc13 and WT ubiquitin. CHFR cysteine rich domain (CRD) and no-E3 negative controls (lanes 1 and 2, respectively) were included as negative controls for E3 RING specific polyubiquitin chain formatin. Untagged (purchased) and His6-tagged (recombinantly expressed) ubiquitin was included as a control for ubiquitin tag specificity against polyubiquitin chain formation. Polyubiquitin chains were detected by western blot analysis (anti-ubiquitin). CRD= cysteine rich domain, MT= mutant, FL=full-length, His6= histidine (6) tag, Ub = ubiquitin.

Targeted RING domain mutations responsible for ablating FL-CHFR- mediated polyubiquitin chain formation were then mapped back onto the original CHFR RING: Ubc13 model derived from superimposing the E3 homology model of the RNF8 protein (PDB: 4ORH, Campbell *et al.*, 2012; Figure 4.6).

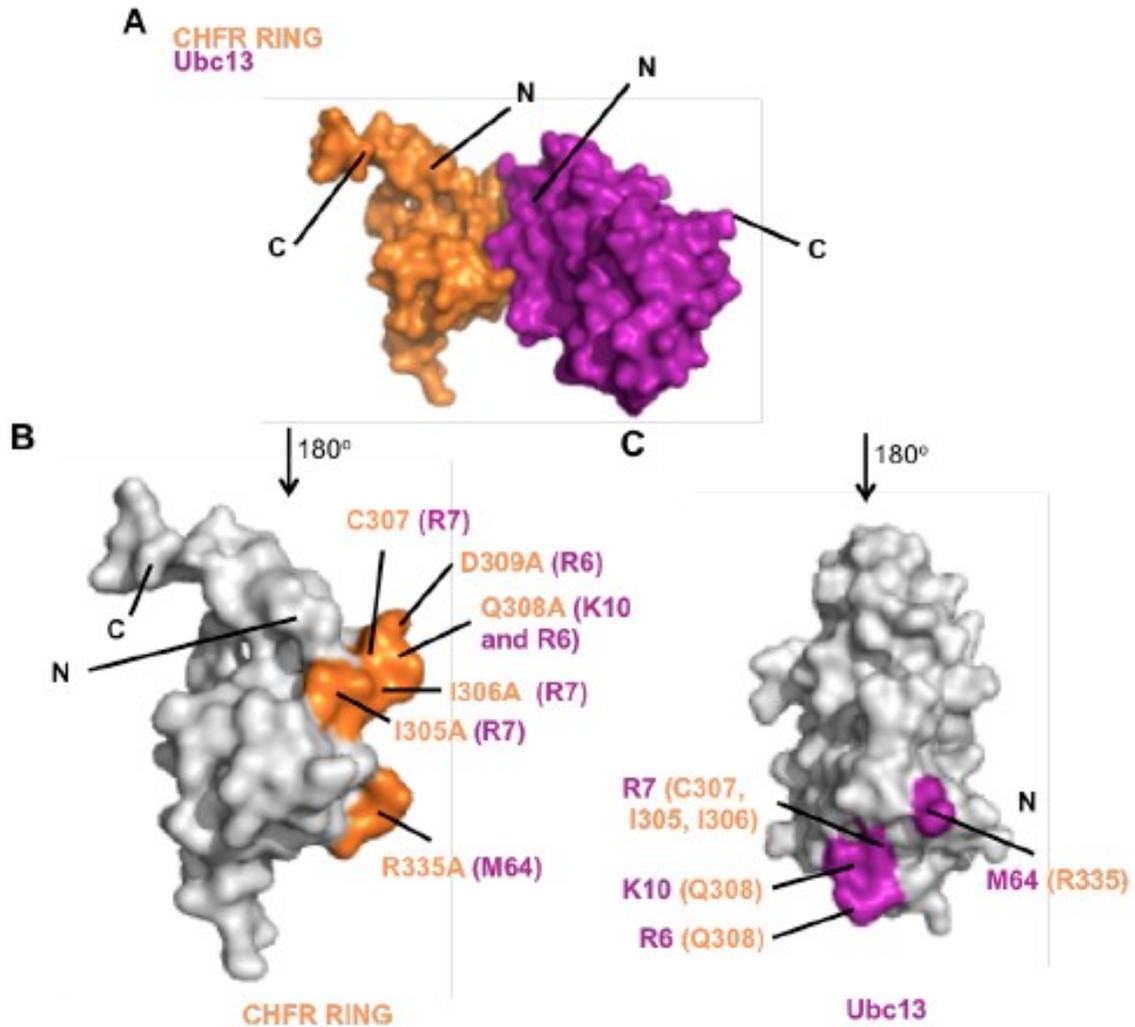


Figure 4.6 Evaluation of CHFR RING residues, modelled to interact with the Ubc13 protein. (A) The CHFR RING model was superimposed onto the RNF8 E3 homologue within the RNF8: Ubc13 experimental structure (Campbell *et al.*, 2012; PDB: 4ORH); with the surfaces of the CHFR RING model (orange) and Ubc13 protein (magenta) indicated. 180° rotation of the (B) CHFR RING model and (C) Ubc13 protein indicates key residues at the CHFR RING (B, orange) and Ubc13 (C, magenta) interface essential in CHFR-mediated polyubiquitin chain formation. For (B) and (C), brackets indicate respective Ubc13 or RING interacting residues, whilst all non-modelled residues are shown in grey.

Utilizing both experimental structures corresponding to RNF8: Ubc13 (PDB: 4ORH) and RNF4: Ubc13 (PDB: 5AIU), it was modelled that respective, homologous residues I305 and I306 within the CHFR RING domain either formed hydrogen bonds with R7 of Ubc13 (PDB: 4ORH) or electrostatic interactions with the Ubc13 SPA motif (96, 97 and 98 respectively) (Figure 4.6). In addition, P340 of the CHFR RING domain was modelled to form electrostatic interactions with the SPA motif of

Ubc13 also. Complete ablation of polyubiquitin chain formation in FL-CHFR RING site directed mutants P340A and I306A is indicative that both residues may be essential for CHFR-mediated polyubiquitin chain formation. In contrast, sufficient residual interactions between FL-CHFR and Ubc13 may be sufficient for the FL-CHFR I305A to still form polyubiquitin chains. In addition, hydrogen bond formation between R335 of the CHFR RING domain and M64 of Ubc13 is likely to be essential in polyubiquitin chain formation, with no chains detected in the presence of the R335A site directed mutant.

Salt bridge and hydrogen bond formation was modelled between CHFR RING residue Q308 and Ubc13 residues R6 and K10 via modelling the CHFR RING domain: Ubc13 interface according to conserved residues with RNF4 (PDB: 5AIU). In addition, hydrogen bond formation between CHFR RING domain residues Q308 and D309 was modelled using the RNF8: Ubc13 experimental structure (PDB: 4ORH). Polyubiquitin chain formation via FL-CHFR RING site directed mutants Q308A and D309A indicate sufficient residual interactions are maintained between the CHFR RING domain and Ubc13 protein.

CHFR RING residue H322 was modelled to stack with a second CHFR RING protein (at Y362), interact with G35 of ubiquitin and also form a hydrogen bond with E34 of ubiquitin. However, ubiquitination assays using individual Y362A and H322C single mutants, in addition to the H322C/ Y362A double mutant, indicate that all component residues within the stacking interaction are essential for the formation of the Ubc13: CHFR RING interface. However, differentiating the importance of hydrogen bond formation with ubiquitin residue E34 or stacking interactions in CHFR-mediated polyubiquitin chain formation is not possible via the CHFR RING domain site directed mutants used. Interestingly, conservation of a non-aromatic, hydrophobic residue is not essential in CHFR mediated polyubiquitin chain formation, as modelled via the CHFR: UbcH5a ~ubiquitin interaction (via PDB: 4AP4) and inclusion within the iterative model (CHFR RING model 10).

Targeted RING domain mutations responsible for ablating CHFR-specific polyubiquitin chain formation were then mapped back onto the original CHFR RING: ubiquitin model derived from superimposing the E3 homology model of the RNF4 protein (PDB: 5AIU, Branigan *et al.*, 2015; Figure 4.7).

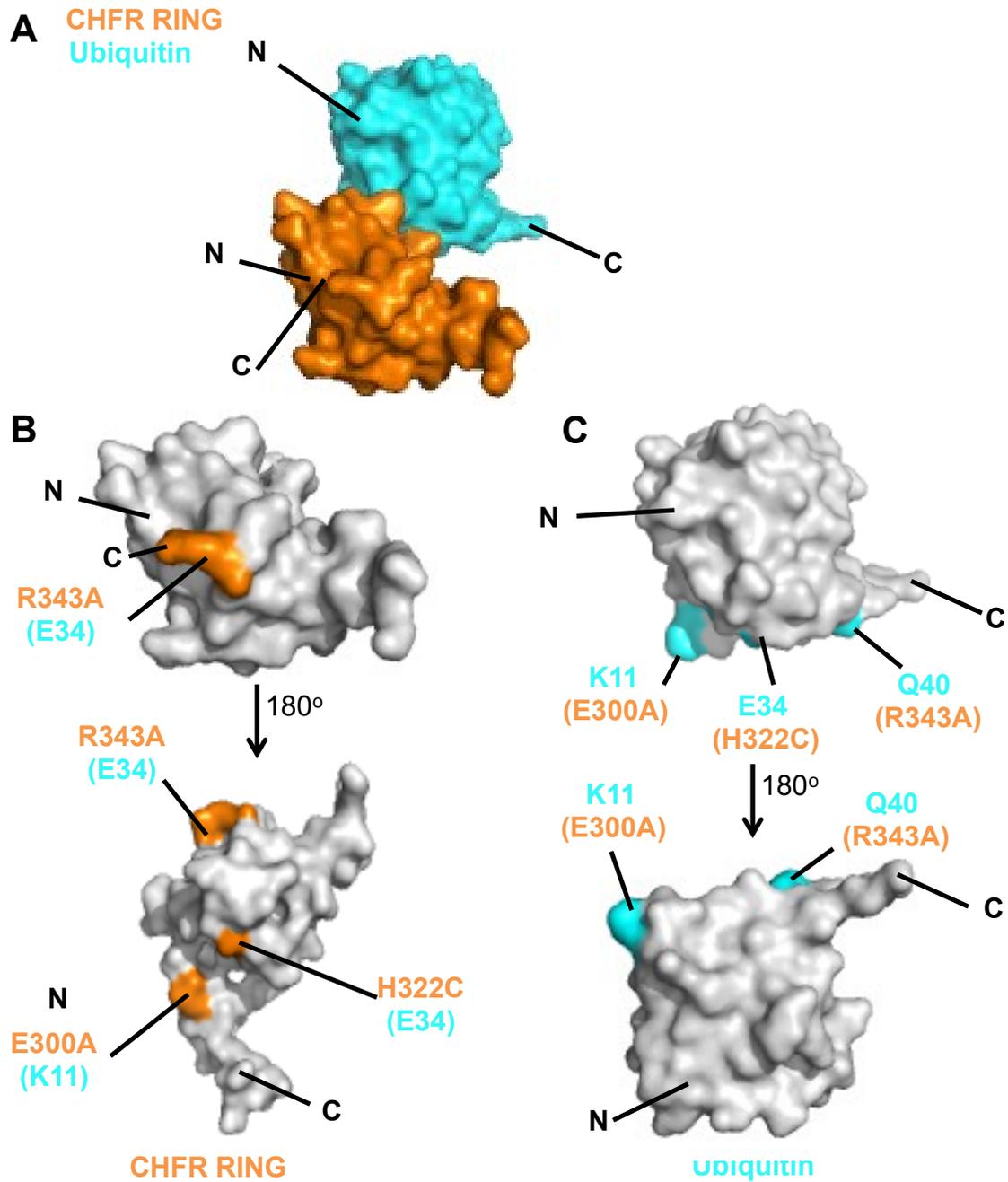


Figure 4.7 Evaluation of CHFR RING residues, modelled to interact with the ubiquitin protein. (A) The CHFR RING model was superimposed onto the RNF4 E3 homologue within the RNF4: Ubc13 ~ubiquitin experimental structure (Branigan *et al.*, 2015; PDB: 5AIU); with the surfaces of the CHFR RING model (orange) and ubiquitin protein (cyan) indicated. 180° rotation of the (B) CHFR RING model and (C) ubiquitin protein indicates key residues at the CHFR RING (B, orange) and ubiquitin (C, cyan) interface essential in CHFR-mediated polyubiquitin chain formation. For (B) and (C), brackets indicate respective ubiquitin or RING interacting residues; whilst all non-modelled residues are shown in grey.

Ubiquitination assays were also performed using (recombinantly expressed) ubiquitin site directed mutants, with WT FL-CHFR and WT Ubc13 proteins (Figure 4.8).

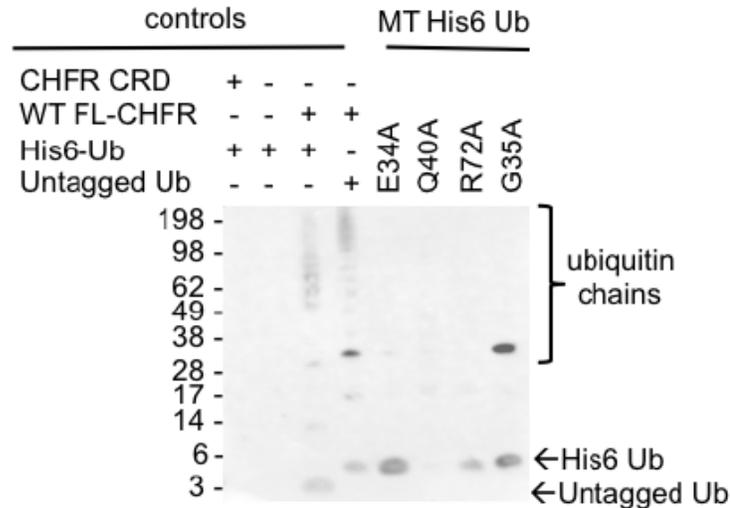


Figure 4.8 Mutational analysis of the FL-CHFR: Ubc13 (:Mms2)~ Ub complex using ubiquitin site directed mutants. Ubiquitination assays were performed using site-directed ubiquitin mutant proteins with WT FL-CHFR and WT Ubc13. CHFR cysteine rich domain (CRD) and no-E3 negative controls (lanes 1 and 2, respectively) were included as negative controls for E3 RING specific polyubiquitin chain formation. Untagged (purchased) and His6-tagged (recombinantly expressed) ubiquitin was included as a control for ubiquitin tag specificity against polyubiquitin chain formation. Polyubiquitin chains were detected by western blot analysis (anti-ubiquitin). CRD= cysteine rich domain, MT= mutant, FL=full-length, His6= histidine (6) tag, Ub = ubiquitin.

Interestingly, ablation of polyubiquitin chain formation in the presence of the E34A ubiquitin site directed mutant specifically indicates hydrogen bond formation with H322 of FL-CHFR is most likely essential in forming the CHFR RING: ubiquitin interface (Figure 4.8); corroborating with observed lack of polyubiquitin chain formation observed in the presence of the FL-CHFR RING H322C site directed mutant. This is independent of the initially modelled stacking interaction of the CHFR RING Y362 residue and observed lack of polyubiquitin chain formation via the H322C/Y362A mutant, since WT ubiquitin was used in the respective assay.

Lack of polyubiquitin chain formation in the presence of the Q40 ubiquitin mutant (Figure 4.5) corroborates with observed ablation of chain formation in the assay utilizing WT ubiquitin and the R343 site directed RING mutant (Figure 4.8). This

strongly indicates hydrogen bond formation between ubiquitin Q40 and CHFR RING R343 is highly important in formation of the CHFR RING: ubiquitin interface and polyubiquitin chain formation. However, the low band intensity corresponding to Q40 on the blot may be indicative of the Q40 site directed mutant's instability within the *in vitro* ubiquitination assay.

Polyubiquitin chain formation in the presence of the G35 ubiquitin site directed mutant indicates a stacking interaction with Y362 and H322 of the CHFR RING domain is likely to be essential in polyubiquitin chain formation (Figure 4.8), corroborating with absence of CHFR-mediated polyubiquitin chain formation observed in assays including FL-CHFR RING site directed mutants H322C, Y362A and H322C/Y362A (Figure 4.5).

4.3 DISCUSSION

Whilst investigating the CHFR interaction with HDAC1 and its targeted degradation via polyubiquitin chain formation, Oh *et al* (2009) produced a FLAG-tagged CHFR RING-domain mutant deficient in E3 ubiquitin ligase activities (Flag-CHFR^{I306A}); indicating an exceptionally important role for residue I306 in maintaining the CHFR RING domain functional capacity. In addition, inhibition of CHFR-mediated polyubiquitin chain formation observed in the same mutant (CHFR^{I306A}) in the presence of the UbcH5a E2 ubiquitin conjugating enzyme was previously observed (Bae *et al.*, 2013); further indicating that I306 is an important site for forming both CHFR RING: Ubc13 and CHFR RING: UbcH5a interfaces.

To model residues at the E3: E2 interface, the CHFR RING model 10 was aligned with the RNF8 experimental structure (Figure 4.1, Panel B; Campbell *et al.*, 2012, PDB: 4ORH); with predicted corresponding residues extrapolated from the RNF8: CHFR RING multiple sequence alignment (Figure 4.1, Panel A; Campbell *et al.*, 2012, PDB: 4ORH). Electrostatic interactions are therefore likely to form between the CHFR RING domain side chain residues P304, I305 and I306 (located within the ZNF1 loop) and the SPA (96, 97, 98) motif of Ubc13; with significant sequence

conservation and identical residue orientation observed with the RNF8 experimental X-ray structure (Campbell *et al.*, 2012, PDB: 4ORH). In contrast, CHFR RING residue T341, corresponding to RNF8 I439, is unlikely to participate in this interaction, considering the difference in residue physiochemical properties.

Interestingly, whilst RNF8 I404D, I405A and I439D mutations ablate RNF4: Ubc13 complex formation (assessed via size exclusion chromatography, SEC), sufficient residual interactions between the E2 and RNF8 mutant proteins (I404D and I405A) enable the E3 ubiquitin ligase to catalyse Lys63 poly-ubiquitin chain formation (Campbell *et al.*, 2012). This suggests both RNF8 (Campbell *et al.*, 2012) and the CHFR RING domain homologous residues (I305 and I306) do not equally contribute to the E2: E3 interaction with the Ubc13 conserved SPA motif; with ablation of CHFR-specific polyubiquitin chain formation identified within assays containing the CHFR I306A mutant, but not those containing the I305A counterpart.

Other mutations documented also suggest additional interactions may be important in maintaining the CHFR: E2 interface. Whilst mutation of the tryptophan to alanine at position 322 resulted in ablation of CHFR E3 ubiquitination ligase activities, a semi-conservative mutation maintained some of the RING domain activity, albeit lower than the wild type counterpart (Bothos *et al.*, 2003). This corroborates strongly with findings regarding the Cbl E3 ubiquitin ligase (Lovick *et al.*, 1999) and this present study.

In addition, loss of CHFR mitotic checkpoint activity was verified within a U2OS cell line, transformed with a GFP-tagged CHFR RING domain mutant (I306A); whereby chromosomal condensation was evaluated 16 hours post-nocodazole (microtubule poison) administration (Bothos *et al.*, 2003). U2OS cells with ectopically expressed wild type CHFR exhibited a lower mitotic index (thus, lower fraction of cells with condensed chromosomes) in comparison the vector control, whilst mutant cells

exhibited both loss of ubiquitin ligase activities and mitotic checkpoint (Bothos *et al.*, 2003).

Bothos *et al* (2003) previously demonstrated ablation of the CHFR: Ubc13 interaction and polyubiquitin chain formation via a H332Q mutant. However, considering this residue coordinates with zinc at the second metal ion centre (Zn²⁺) alongside C342, C320 and C339; this is most likely to be resultant of destabilization of the loop within this region, rather than specific targeting of residues at the E2: E3 or E2: ubiquitin interfaces. In contrast, H322C mutants included within this study retain metal zinc ion co-ordination, indicative that the proposed Y362: H322 stacking interaction (with G35 of ubiquitin) may play an important role in CHFR- mediated polyubiquitin chain formation (Figures 4.5 and 4.8), as evident in the RNF4 (E3 homologue) co-crystal complex with ubiquitin and Ubc13 (Branigan *et al.*, 2012) and RNF4: Ubch5a ~ubiquitin complex (Plechanovova *et al.*, 2012).

Whilst only the CHFR interaction with Ubc13 and Ubch5a were modelled in the present study, applicability of current findings in modelling other CHFR RING: E2 interfaces are highly probable. For example, Kang *et al* (2002) demonstrated two point mutations within the CHFR RING domain described here (I306A, W322A) also ablate CHFR-mediated polyubiquitin chain formation in the presence of Ubch4. However, thioester conjugate assays of mutant ubiquitin proteins may be required to verify specific disruption of the RING: ubiquitin interface and not simply an ablation of the E2's capacity to form a thioester linkage with the ubiquitin site directed mutant. In addition, site directed Ubc13 mutants designed to target and disrupt key interactions at the CHFR RING: Ubc13 interface could further support the conclusions presented here.

CHAPTER 5: Conclusions

E3 ubiquitin ligases play an incredibly important role in the regulation of numerous intracellular processes, including mitotic cell division. Frequent down-regulation of CHFR (via promoter hypermethylation or mutation) has been identified in numerous cancer cell lines and tumours and associated with loss of the antepause checkpoint (Pines and Matsusaka, 2004, Scolnick and Halazonetis, 2002).

Through exploring the biophysical, structural and biochemical properties of the central CHFR RING domain, greater insight into the structural basis of CHFR-mediated polyubiquitin chain formation has been obtained.

Firstly, the FL-CHFR protein is dimerized via by the N-terminal FHA domain, verified using SEC-SAXS and SEC experiments, in corroboration with the domain-swapped FHA domain dimer X-ray structure (Stravridi *et al.*, 2012). In addition, dimerization of Δ FHA CHFR proteins (verified using SEC with MWS) and multimer formation of the C-terminal CRD (via SEC with MWS and native page) indicate that the FL-CHFR may also dimerize via the CRD. However, Scolnick and Halazonetis (2002) identified the dominant negative effects of CRD and FHA domain deletion mutants ablate the capacity of endogenous FL-CHFR to delay mitotic progression in the presence of microtubule stress. Whilst this dimerization is not essential for Δ FHA- Δ CRD-CHFR (Kang *et al.*, 2002) or Δ FHA-CHFR mediated polyubiquitin chain formation (as demonstrated in this study), dimerization of FL-CHFR via the FHA or CRD may play an incredibly important role in recognition of other antepause checkpoint protein components, essential in delaying mitotic progression.

Dimerization of the FL-CHFR protein strongly corroborates with the modelled H322: Y362 CHFR RING domain stacking interaction, suggesting at least two CHFR RING domain contributions of either aromatic/ hydrophobic residue is essential in forming a stable interface with Ubc13 and to facilitate polyubiquitin chain formation. However, Kang *et al* (2002) have identified Δ FHA- Δ CRD-CHFR retains the capacity to form polyubiquitin chains *in vitro* in the presence of the Ubc4

E2 ubiquitin conjugating enzyme. The specific stacking interaction and dimerization modelled using FL-CHFR and Δ FHA-CHFR within this study may be specific to the CHFR RING interaction with the Ubc13: Mms2 E2 heterodimer and ubiquitin.

The current FL-CHFR working model therefore postulates that all three CHFR domains have the capacity to dimerize in solution, whereby dimerization may be required for its E3 ubiquitin ligase activities in the presence of the Ubc13: Mms2 heterodimer but not the Ubc4 E2 ubiquitin conjugation enzyme (Kang *et al.*, 2002) (Figure 5.1).

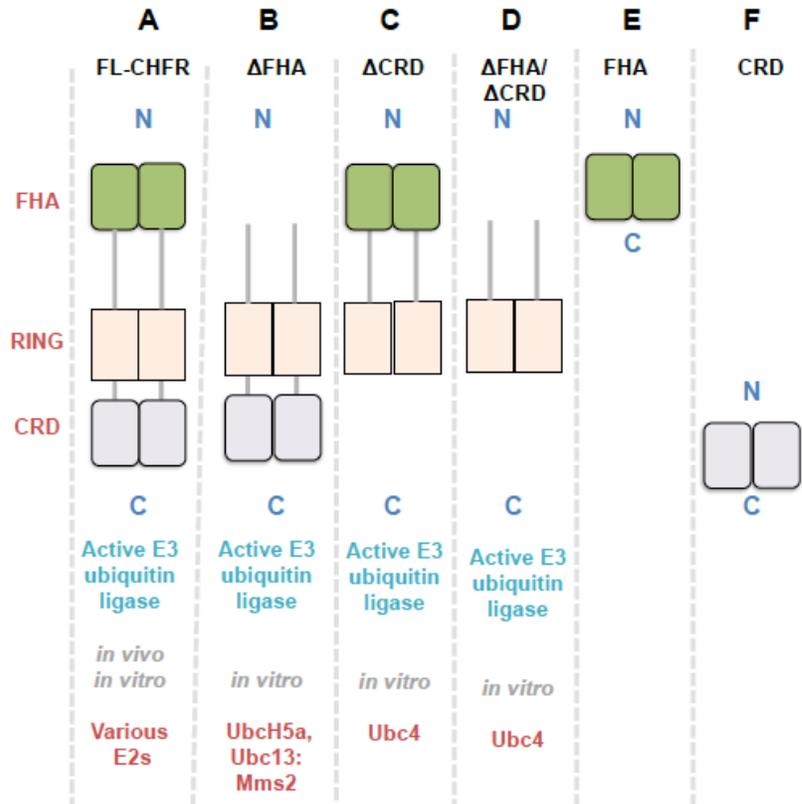


Figure 5.1 Model for the dimerization of CHFR. (A) Within the current model, full-length CHFR protein dimerizes within all three domains (N-terminal FHA, central RING and C-terminal cysteine rich (CRD)) and is an active E3 ubiquitin ligase via *in vitro* and *in vivo* polyubiquitin chain formatin (Scolnick and Halazonetis, 2002; Kang *et al.*, 2002). (B) Upon deletion of the N-terminal FHA domain (Δ FHA), CHFR remains dimeric (via dimerization at the C-terminal CRD and possibly the central RING domains); retaining its E3 ubiquitin ligase activity *in vitro* in the presence of UbcH5a and Ubc13: Mms2 E2 enzymes (as discussed in this study). However, the dominant negative effects of the protein result in ablation of endogenous FL-CHFR's capacity to delay mitotic progression at the antepause checkpoint in response to mitotic stress *in vivo* (Scolnick and Halazonetis, 2002). (C) Kang *et al* (2002) have demonstrated the Δ CRD-CHFR is an active E3 in *in vitro* (with Ubc4), as is the (D) Δ FHA- Δ CRD-CHFR double-domain mutant, retaining the capacity to form polyubiquitin chains in the presence of Ubc4 *in vitro*. (E) The CHFR FHA domain dimerizes in solution (Stravidi *et al.*, 2002, also verified in this study) whilst the (F) CRD form dimers and higher order oligomeric states in solution (this study).

Utilization of an E2Scan ubiquitination assay has facilitated the identification of different E2 ubiquitin ligases responsible for CHFR-mediated polyubiquitin chain formation *in vitro*. The wide range of different lysine linkages between ubiquitin proteins within the chains formed by these verified E2s (K11, -48 and -63; Chapter 1, Table 1.1) exemplify the diverse intracellular roles of the CHFR protein in cell cycle regulation, DNA repair, NF κ B signalling and targeting proteins to the 26S

proteasome. Verification of CHFR's ubiquitin ligase activities with the E2s *in vitro* and *in vivo* would be required to verify and further characterize these enzymatic reactions within a cellular environment.

Production of a RING domain homology model, combined with detailed bioinformatical analysis of RING E3 homologues within E3: E2: ubiquitin complexes, proved exceptionally useful in modelling the key CHFR RING: Ubc13 and CHFR RING: ubiquitin interactions essential for CHFR-mediated polyubiquitin chain formation *in vitro*. 10 out of the 15 site directed CHFR RING mutants were incapable of K63-linked polyubiquitin chain formation *in vitro*, with disruption of all 4 ubiquitin-targeted interactions with the CHFR RING (model) successfully ablating polyubiquitin chain formation.

The present work has both verified and broadened knowledge concerning the oligomeric states of individual CHFR domains and deletion mutants, the role of CHFR oligomerization in polyubiquitin chain formation, the oligomeric state of the FL-CHFR proteins and E2 ubiquitin conjugating enzymes responsible for CHFR polyubiquitin chain formation. In addition, a wide range of interactions at the CHFR RING: E2 and CHFR RING: ubiquitin interfaces have been successfully modelled using a homology model of the central RING domain, with interactions and consequence upon polyubiquitin chain formation verified *in vitro*.

In order to elucidate the mechanisms underpinning interactions between the CHFR RING domain and ubiquitination pathway components (both substrate and enzymatic), an experimental structure of the FL-CHFR protein in complex with a thioester conjugated E2 and ubiquitin (E2~ ubiquitin) could be obtained by X-ray crystallography or cryo-EM. This would provide an invaluable insight into the molecular mechanisms underpinning CHFR's interactions with ubiquitin and E2 ubiquitin conjugating enzymes within the ubiquitination enzymatic cascade; prior to ubiquitination of mitotic substrates and other important protein-protein interactions that are essential in mediating the antephasic checkpoint.

CHAPTER 6: Appendices

Appendix 1	224
Appendix 2	257
Appendix 3	271

Appendix 1

A1.1 CHEMICALS

Analytical grade chemicals ($\geq 98\%$, HPLC) were used throughout (Table A1.1).

Table A1.1 Chemical, reagent and buffer suppliers

Reagent	Supplier
2-(4-Hydroxyphenylazo)-benzoic acid (HABA)	Sigma
β -mercaptoethanol	Sigma
Bromophenol blue	Sigma
Chloramphenicol	Sigma
d-Desthiobiotin	Sigma
DTT	Melford
EDTA	Sigma
Ethanol	Sigma
Glycerol	Melford
Glycine	Fisher Scientific
HEPES	Sigma
Hydrogen chloride solution	Sigma
Imidazole	Acros Organics
IPTG	Melford
Kanamycin sulfate	Sigma
MES SDS running buffer (20X)	ThermoFisher Scientific
Methanol	Fisher Scientific
Nickel (II) Chloride	Sigma
PBS	Melford
PMSF	Melford
Ponceau S red staining solution	Sigma
Sodium chloride	Fisher Scientific

Table A1.1 Chemical, reagent and buffer suppliers (...Continued)

Reagent	Supplier
Sodium hydroxide solution (5M)	Fisher Scientific
Sodium phosphate	Acros Organics
SDS micropellets	Fisher Scientific
Skimmed milk powder (Marvel)	Premier Foods
Spectinomycin dihydrochloride pentahydrate	Melford
PBS	Melford
Ponceau S	Sigma
Tris base	Sigma
Tween 20	Sigma
Zinc chloride	Sigma

Western blot transfer buffer (25mM Tris Base, 192mM glycine, 20% v/v methanol) was prepared by Dr Mark Richards and Mrs Honglin Rong (Astbury Centre for Structural Molecular Biology, University of Leeds).

His6-tagged recombinant TEV protease (1 mg/mL in 5% v/v glycerol) was expressed, purified and flash-frozen by Dr Mark Richards (Astbury Centre for Structural Molecular Biology, University of Leeds).

A1.2 BIOCHEMICALS

Expression competent BL21 CodonPlus (DE3) RIL cells (Stratagene) were prepared by Dr Mark Richards (Astbury Centre for Structural Molecular Biology, University of Leeds). LB-agar plates and lysogeny-Broth (LB) were produced by Dr Mark Richards and Mrs Honglin Rong (Astbury Centre for Structural Molecular Biology, University of Leeds).

A1.3 VECTORS

Vectors used for protein expression were supplied by various companies and constructs produced by Dr Mark Richards (Astbury Centre for Structural Molecular Biology, University of Leeds, Table A1.2).

Table A1.2 Bacterial protein expression constructs

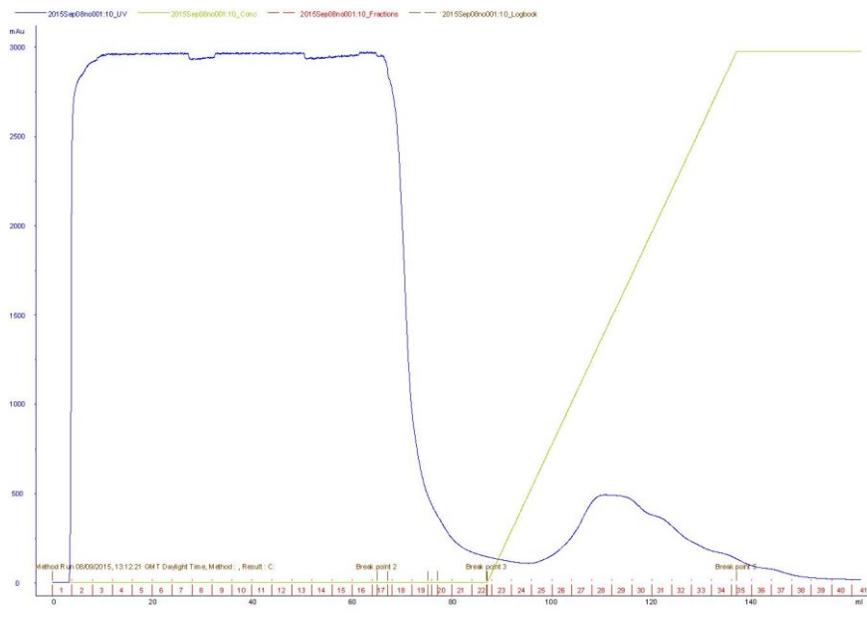
Protein	Vector	Tags, linkers and cleavage sites	Supplier	Antibiotic resistance
WT FL-CHFR (isoform 2, 1-664)	pRSF(Duet)	N-terminal tags: His-FLAG-Strep2- linker-Strep2- PreScission site	Novagen	Kan
CHFR Δ FHA-1 (isoform 2, 231-664)	pRSF(Duet)	N-terminal tags: His-FLAG-Strep2- linker-Strep2- PreScission site	Novagen	Kan
CHFR Δ FHA-2 (isoform 2, 251-664)	pRSF(Duet)	N-terminal tags: His-FLAG-Strep2- linker-Strep2- PreScission site	Novagen	Kan
CHFR Δ FHA-3 (isoform 2, 267-664)	pRSF(Duet)	N-terminal tags: His-FLAG-Strep2- linker-Strep2- PreScission site	Novagen	Kan
CHFR C-terminus (C1, 394-664)	pETM6T1	N-terminal tag: TEV-cleavable His ₆ -NusA tag	Novagen (derived from pET44)	Kan

Table A1.2 Bacterial protein expression constructs (...Continued)

Protein	Vector	Tags, linkers and cleavage sites	Supplier	Antibiotic resistance
CHFR FHA (13-180)	pET-30(TEV)	N-terminal, TEV- cleavable His-tag	Novagen	Kan
UbcH5a	pET-30(TEV)	N-terminal, TEV- cleavable His-tag	Novagen	Kan
Ubc13	pET-30(TEV)	N-terminal, TEV- cleavable His-tag	Novagen	Kan
Mms2	pET-30(TEV)	N-terminal, TEV- cleavable His-tag	Novagen	Kan
Ubiquitin	pET17	N-terminal, 3C- cleavable His-tag	Novagen	Amp

A1.4 PURIFICATION OF CHFR PROTEINS

A



B

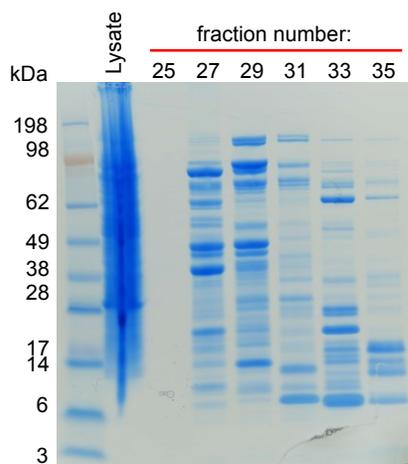


Figure A1.1 Purification of FL-CHFR by ion exchange chromatography. (A) Ion exchange chromatography was performed by loading clarified lysate onto a HiTrap Heparin column (Panel B, Lane 1), with the chromatogram indicating an elution peak (UV, A_{280} ; navy line); corresponding to a linear gradient of NaCl from 200 mM to 1 M (concentration of high-salt buffer B; green line). (B) SDS-PAGE analysis of alternating fractions taken from across the peak (see red, horizontal line in (A)) indicated FL-CHFR potentially at 79.4 kDa, with numerous other proteins also identified.

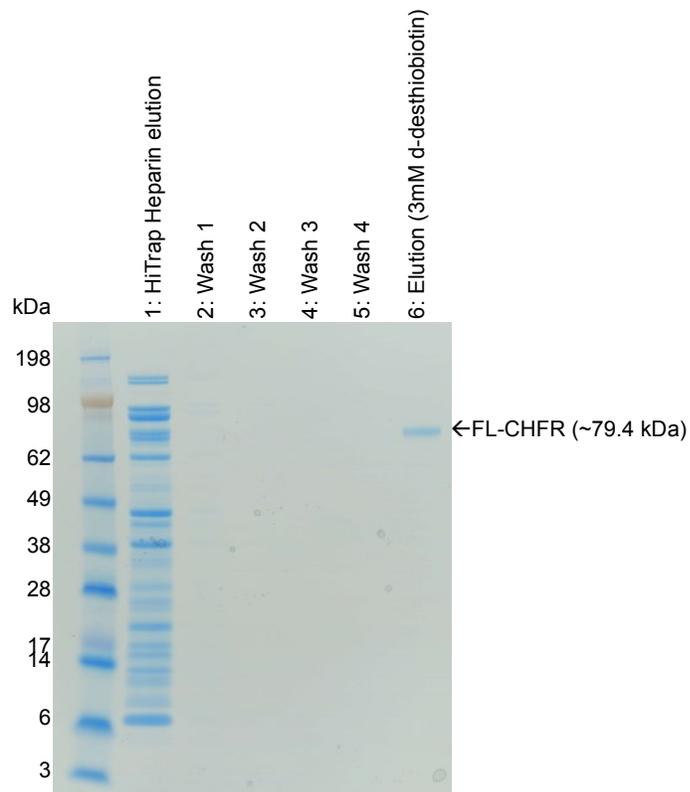


Figure A1.2 Purification of FL-CHFR by affinity chromatography. The original Hitrap Heparin (ion exchange) elution (Lane 1; also see A2.1) consisted of numerous other protein contaminants, in addition to the FL-CHFR protein. Proceeding initial manual injection of the elution onto a *Strep-tactin* column (four times) and four washes using the original lysis (A) Buffer (Lanes 2 to 5); the FL-CHFR protein (~79.4 kDa) was eluted using an elution buffer (B, buffer A supplemented with 3mM d-Desthiobiotin; Lane 6).

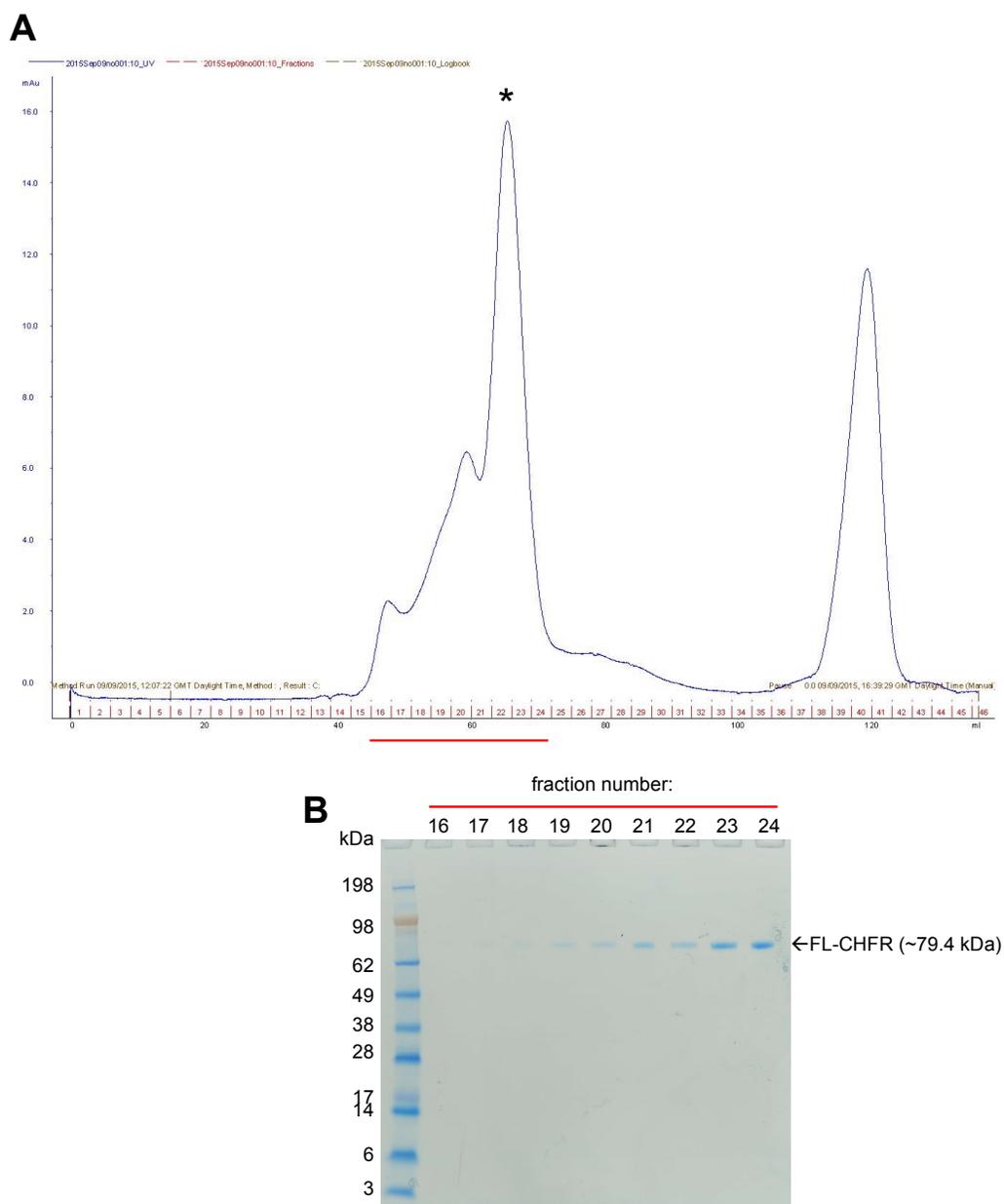


Figure A1.3 Purification of FL-CHFR by size-exclusion chromatography. (A) The FL-CHFR protein was purified using a Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare), with the elution profile (UV, A_{280} nm) indicated in the chromatogram (navy line). (B) Selected fractions (16 to 24, see horizontal red line in chromatogram (A) and peak indicated by an asterisk, *) were then analyzed by SDS-PAGE.

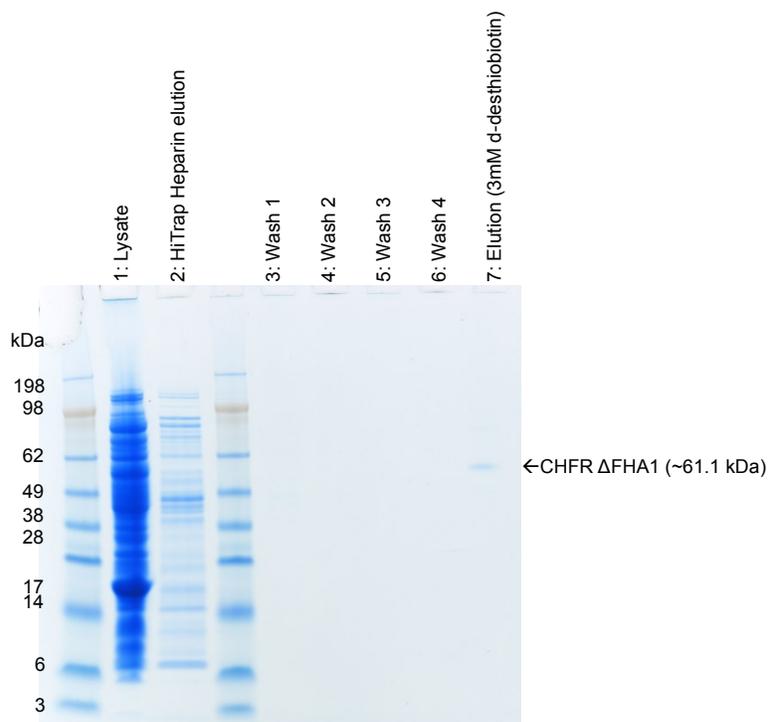


Figure A1.4 Purification of CHFR Δ FHA1 (231-664) by ion exchange and affinity chromatography. Ion exchange chromatography was performed by loading clarified lysate onto a HiTrap Heparin column (Lane 1) and washing unbound proteins prior to elution (Lane 2) with a high salt (1M NaCl) buffer. The original HiTrap Heparin (ion exchange) elution (Lane 2) consisted of numerous other protein contaminants, in addition to the Δ FHA1 protein. Proceeding initial manual injection of the elution onto a *Strep-tactin* column (four times) and four washes using the original lysis (A) Buffer (Lanes 3 to 6); the Δ FHA1 protein was eluted using an elution buffer (B, buffer A supplemented with 3mM d-Desthiobiotin; Lane 7).

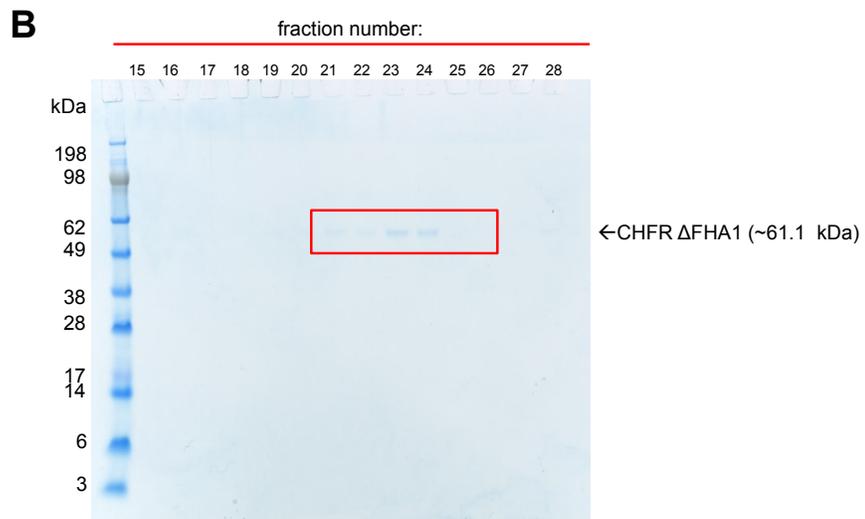
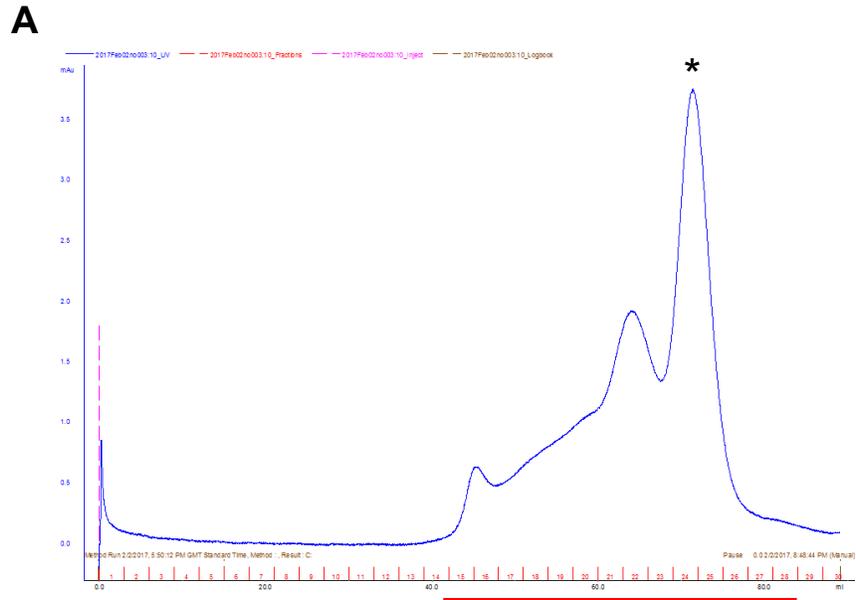


Figure A1.5 Purification of CHFR ΔFHA1 (231-664) by size-exclusion chromatography. (A) The ΔFHA1 protein was purified using a Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare), with the elution profile (UV, A₂₈₀ nm) indicated in the chromatogram (navy line). (B) Selected fractions (19 to 27, see horizontal red line in chromatogram (A) and peak indicated by an asterisk, *) were then analyzed by SDS-PAGE.

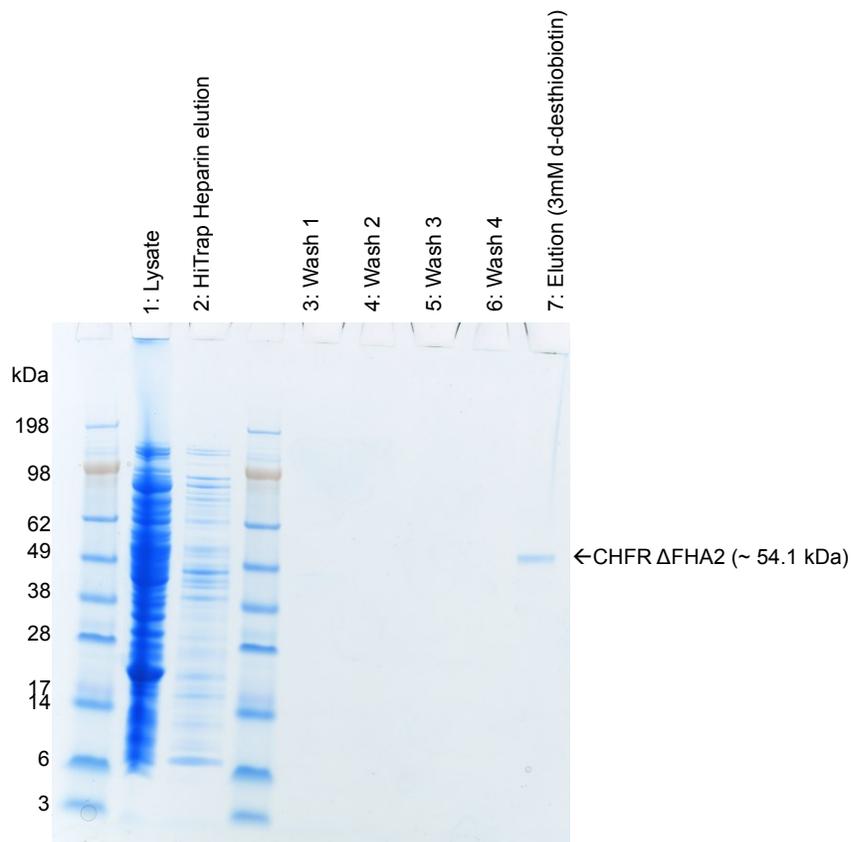


Figure A1.6 Purification of CHFR Δ FHA2 (251-664) by ion exchange and affinity chromatography. Ion exchange chromatography was performed by loading clarified lysate onto a HiTrap Heparin column (Lane 1) and washing unbound proteins prior to elution (Lane 2) with a high salt (1M NaCl) buffer. The original HiTrap Heparin (ion exchange) elution (Lane 2) consisted of numerous other protein contaminants, in addition to the Δ FHA2 protein. Proceeding initial manual injection of the elution onto a Strep-*tactin* column (four times) and four washes using the original lysis (A) Buffer (Lanes 3 to 6); the Δ FHA2 protein was eluted using an elution buffer (B, buffer A supplemented with 3mM d-Desthiobiotin; Lane 7).

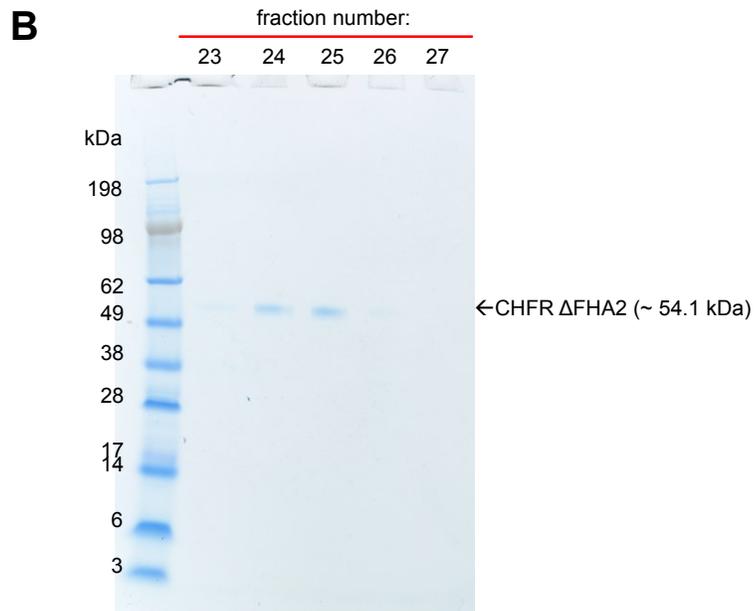
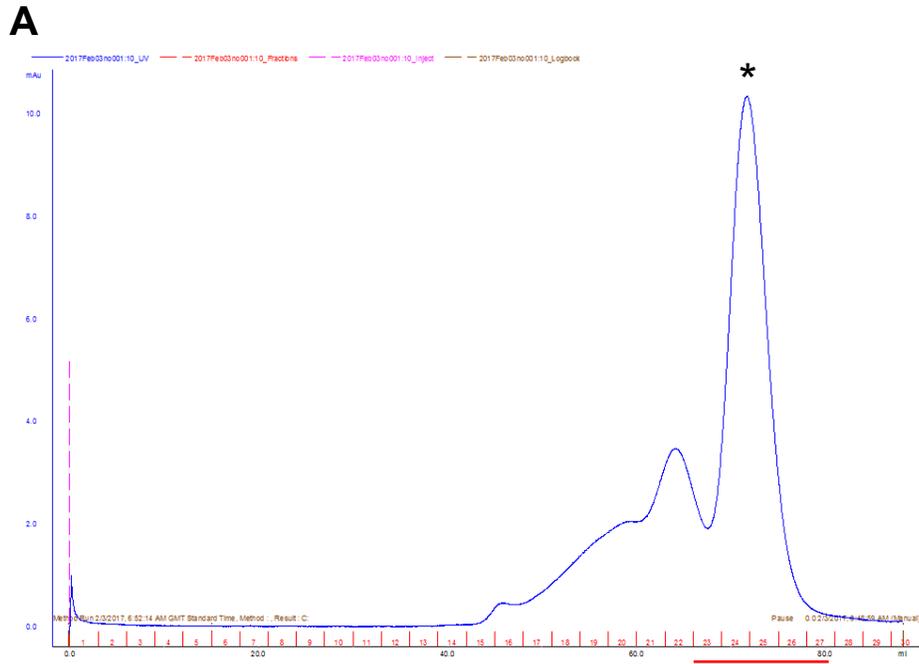


Figure A1.7 Purification of CHFR ΔFHA2 (251-664) by size-exclusion chromatography. (A) The ΔFHA2 protein was purified using a Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare), with the elution profile (UV, A_{280} nm) indicated in the chromatogram (navy line). (B) Selected fractions (23 to 27, see horizontal red line in chromatogram (A) and peak indicated by an asterisk, *) were then analyzed by SDS-PAGE.

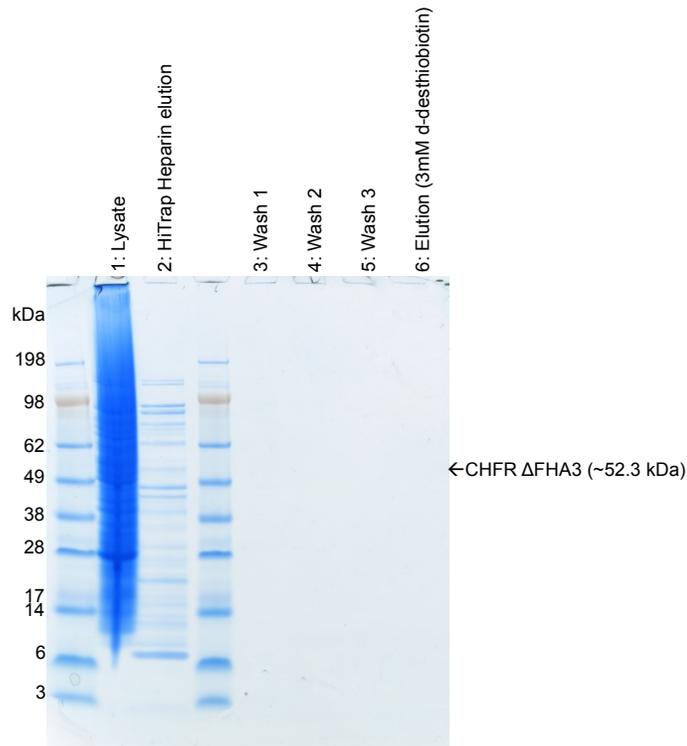


Figure A1.8 Purification of CHFR Δ FHA3 (267-664) by ion exchange and affinity chromatography. Ion exchange chromatography was performed by loading clarified lysate onto a HiTrap Heparin column (Lane 1) and washing unbound proteins prior to elution (Lane 2) with a high salt (1M NaCl) buffer. The original HiTrap Heparin (ion exchange) elution (Lane 2) consisted of numerous other protein contaminants, in addition to the Δ FHA3 protein. Proceeding initial manual injection of the elution onto a *Strep-tactin* column (four times) and four washes using the original lysis (A) Buffer (Lanes 3 to 6); the Δ FHA3 protein was eluted using an elution buffer (B, buffer A supplemented with 3mM d-Desthiobiotin; Lane 6).

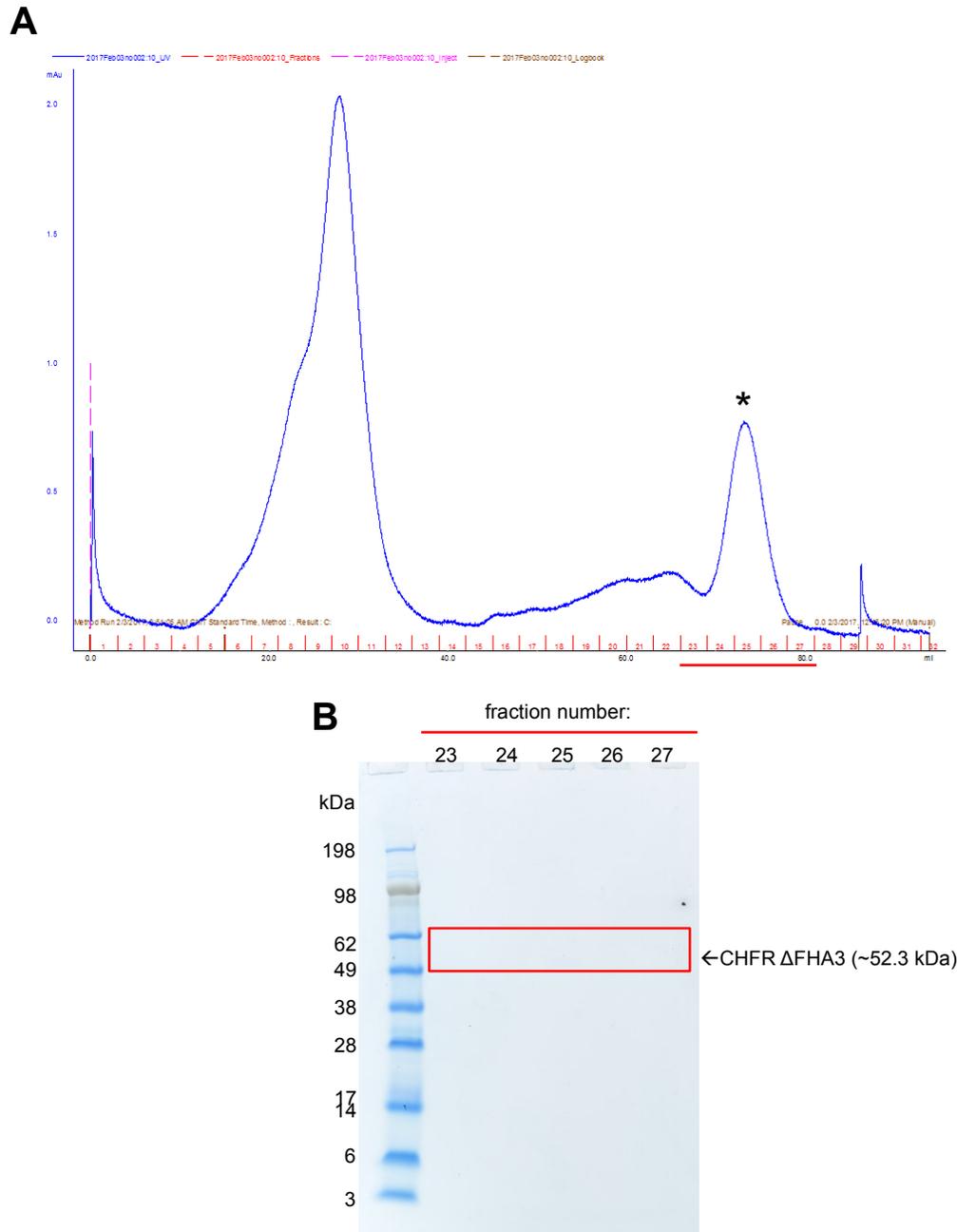


Figure A1.9 Purification of CHFR ΔFHA3 (267-664) by size-exclusion chromatography. (A) The ΔFHA3 protein was purified using a Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare), with the elution profile (UV, A_{280} nm) indicated in the chromatogram (navy line). (B) Selected fractions (23 to 27, see horizontal red line in chromatogram (A) and peak indicated by an asterisk, *) were then analyzed by SDS-PAGE.

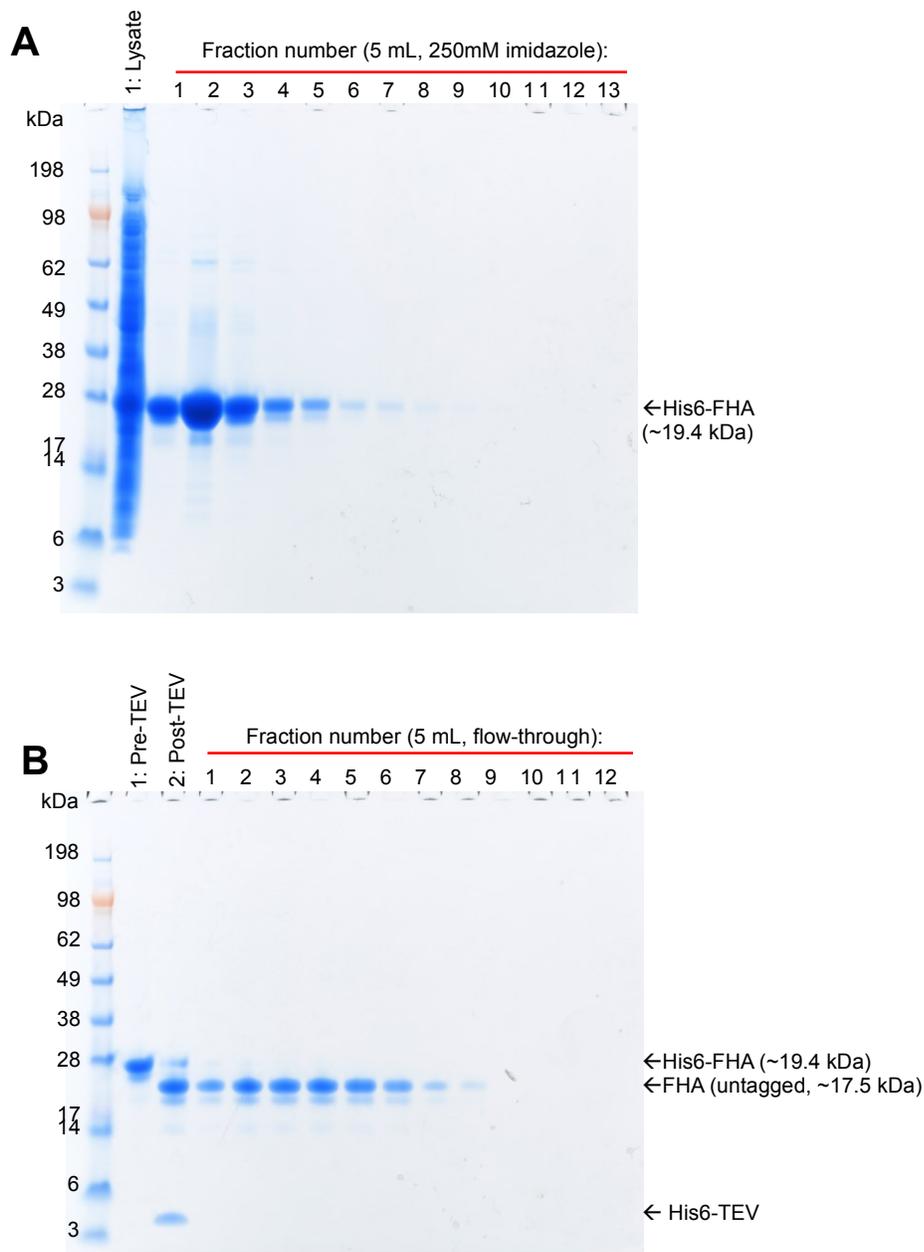


Figure A1.10 Purification of the CHFR FHA domain (13-180) by metal ion affinity chromatography. (A) Proceeding manual injection of the clarified lysate (Lane 1) and washing of the column to remove unbound contaminants; 13 individual (5 mL) elutions of the His6-FHA protein (19.4 kDa) were obtained. Pooled fractions were then dialyzed over 18 hours with TEV protease (B, Lane 1) to cleave the His-tag from the FHA protein; producing a lower molecular weight, untagged FHA domain protein (B, Lane 2, 17.5 kDa). The dialyzed and cleaved protein was then injected for a second time onto a (stripped and re-charged) HisTrap FF column to isolate untagged FHA protein from His-tagged TEV protease and uncleaved (His-tagged) FHA protein.

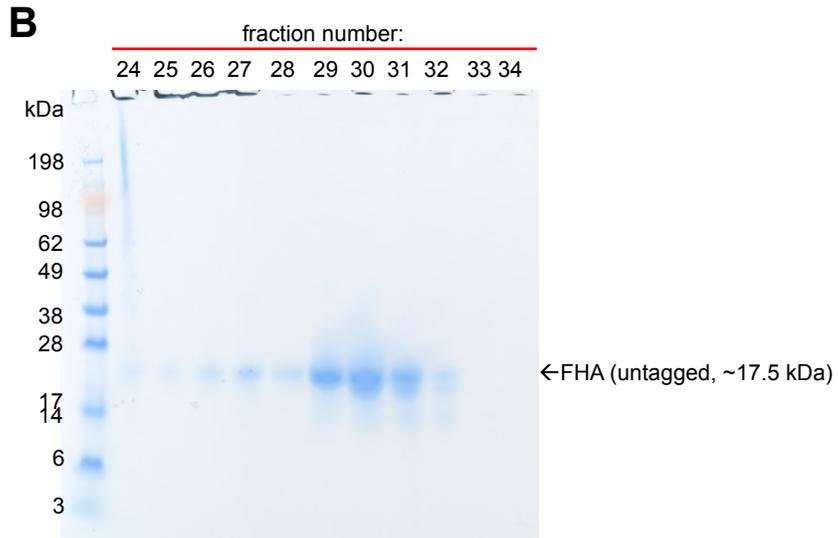
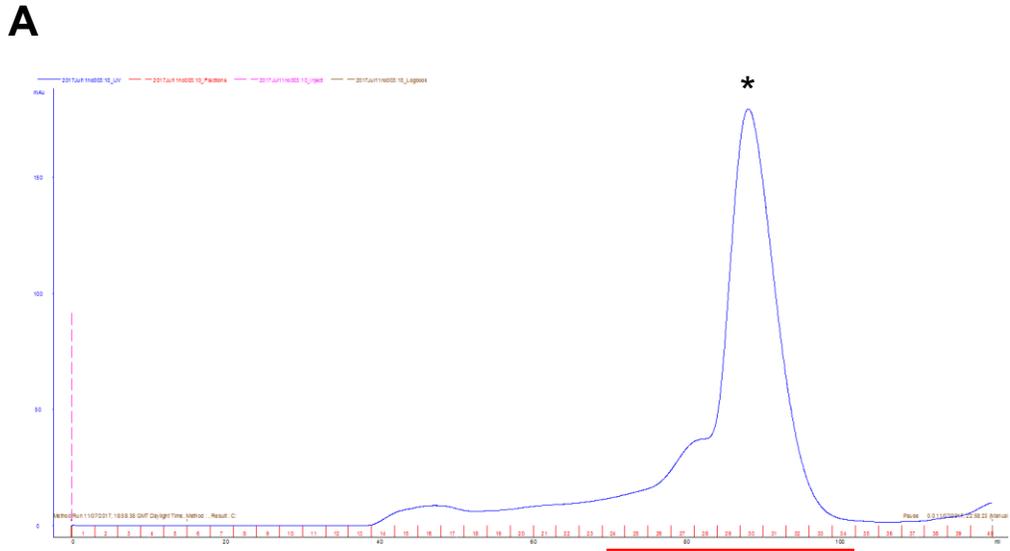


Figure A1.11 Purification of the CHFR FHA domain (13-180) by size-exclusion chromatography. (A) The FHA domain was purified using a Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare), with the elution profile (UV, A_{280} nm) indicated in the chromatogram (navy line). (B) Selected fractions (24 to 34, see horizontal red line in chromatogram (A) and peak indicated by an asterisk, *) were then analyzed by SDS-PAGE.

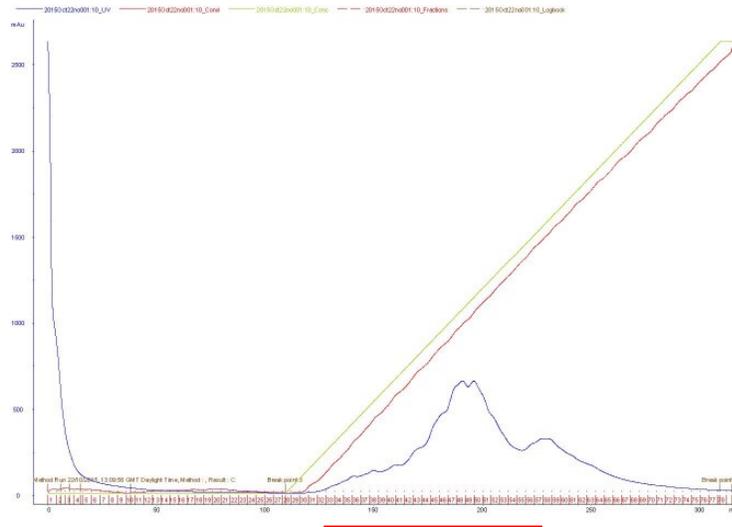
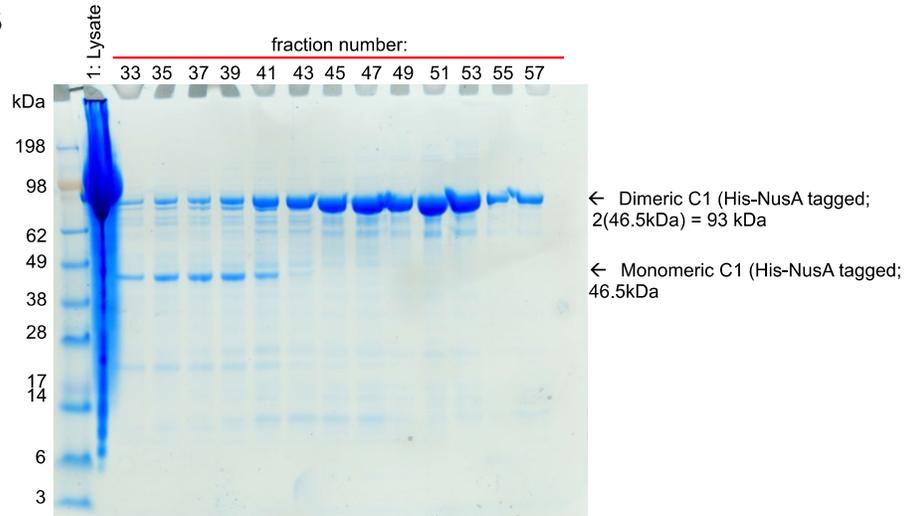
A**B**

Figure A1.12 initial purification of the CHFR C-terminus (394-664) by anion exchange chromatography. (A) Anion exchange chromatography was performed by loading clarified lysate onto a HiTrap ANX FF column (Panel B, Lane 1), with the chromatogram indicating an elution peak (UV, A₂₈₀; navy line); corresponding to a linear gradient of NaCl from 200 mM to 1 M (concentration of high-salt buffer B; green line). (B) SDS-PAGE analysis of alternating fractions taken from across the peak (see red, horizontal line in (A)) indicated the CHFR C-terminus present in both monomeric (46.5 kDa) and dimeric (93 kDa) oligomerization states.

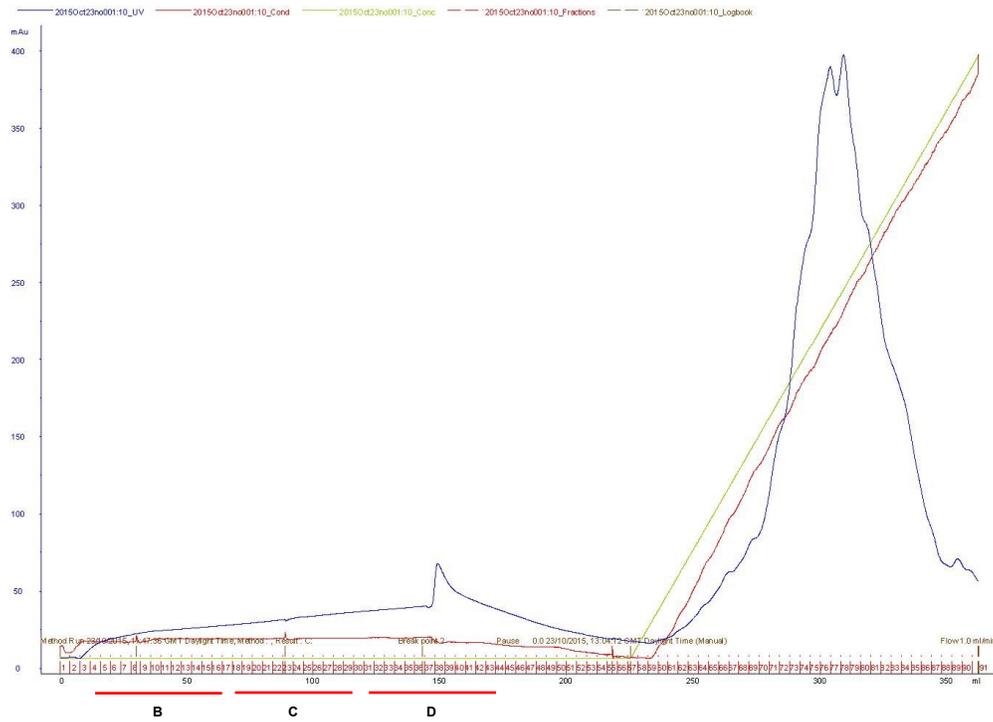
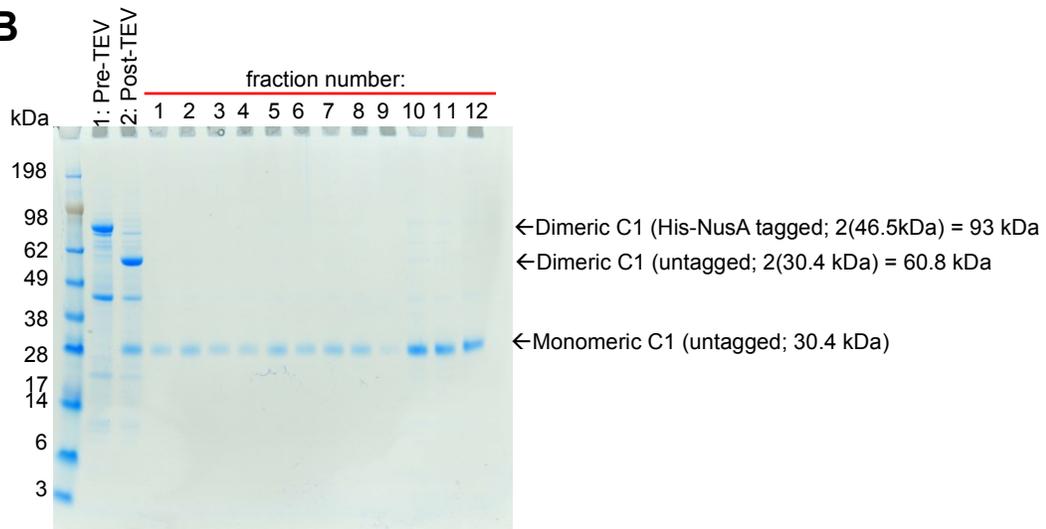
A**B**

Figure A1.13 TEV cleavage of His₆-NusA tag and second anion exchange chromatography of CHFR C-terminus (394-664). Pooled fractions were then dialyzed over 18 hours with TEV protease (Panel B, Lane 1) to cleave the His and NusA tags from the CHFR C-terminus; producing a lower molecular weight, untagged and dimeric C-terminus protein (B, Lane 2, 60.8 kDa). The dialyzed and cleaved protein was then injected for a second time onto a (cleaned) HiTrap ANX FF column to isolate a monomeric (30.4 kDa), untagged C-terminal protein from His-tagged TEV protease and uncleaved (His-NusA tagged) C-terminus protein.

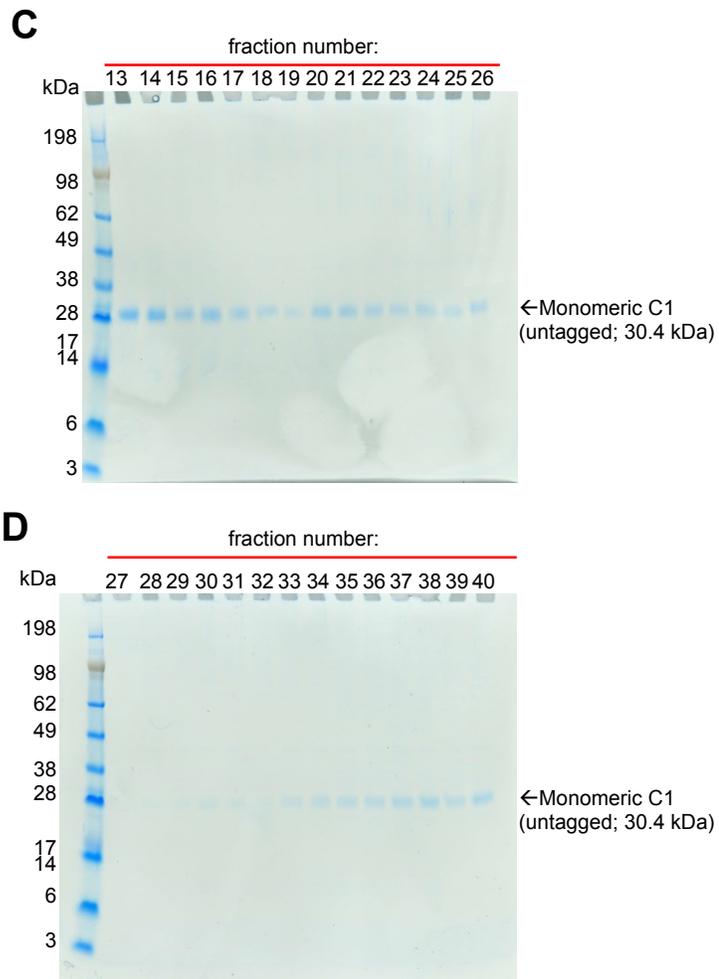


Figure A1.13 TEV cleavage of His6-NusA tag and second anion exchange chromatography of CHFR C-terminus (394-664)....Continued. Panels C and D correspond to fractions denoted in Panel A (see horizontal red line labeled with C and D).

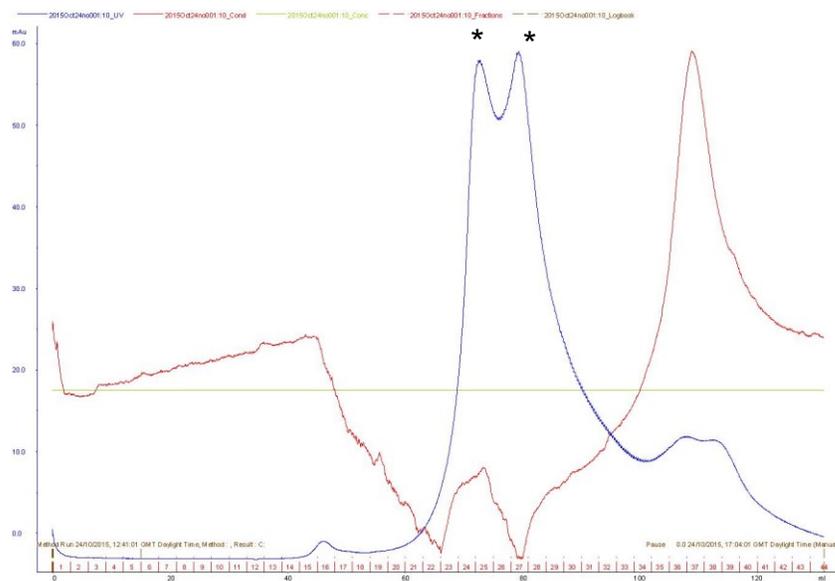
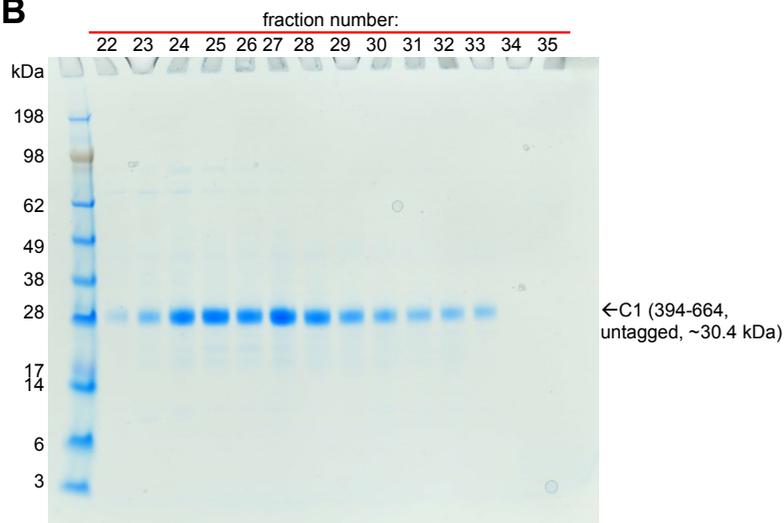
A**B**

Figure A1.14 Purification of the CHFR C-terminus (394-664) by size-exclusion chromatography. (A) The C-terminal cysteine rich domain was purified using a Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare), with the elution profile (UV, A_{280} nm) indicated in the chromatogram (navy line). (B) Selected fractions (22 to 33, see horizontal red line in chromatogram (A) and both peaks indicated by an asterisk, *) were then analyzed by SDS-PAGE.

A1.5 ANALYTICAL SIZE EXCLUSION CHROMATOGRAPHY

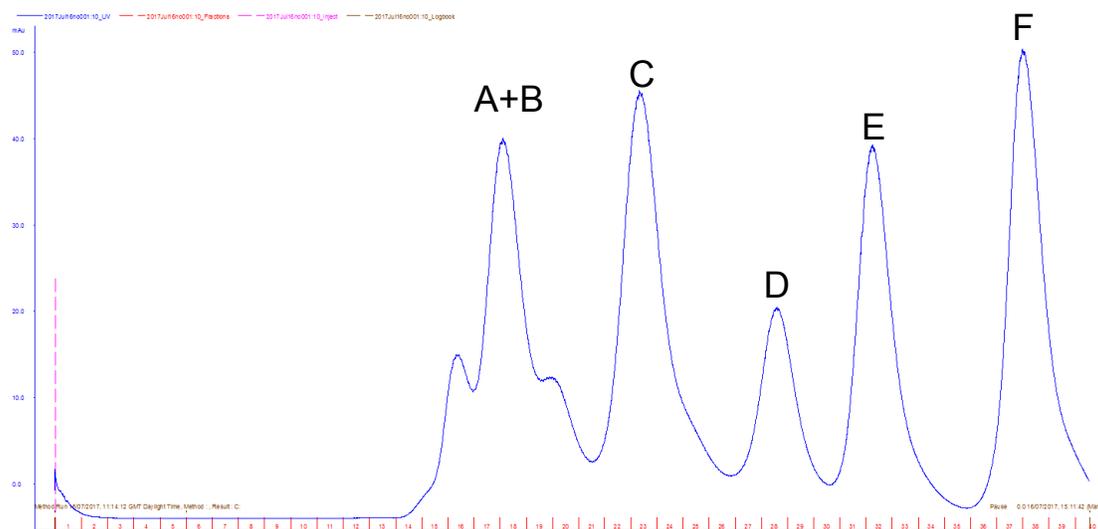


Figure 1.15 Analytical size exclusion chromatography using molecular weight standards. To estimate molecular weights of purified proteins (FL-CHFR, Δ FHA1, Δ FHA2, Δ FHA3, FHA domain and C-terminal cysteine rich domain) a size exclusion chromatography-specific molecular weight standard (BioRad) was analysed using a Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare); with buffer composition identical to purified proteins (150 mM NaCl, 25 mM Tris-HCl (pH 7.0), 2 % (v/v) glycerol). Individual protein elutions, as indicated by the blue (UV) peaks correspond to (A) protein aggregates (void peak) + (B) bovine thyroglobulin (670 kDa; both A and B are eluted as a single group of peaks), (C) bovine γ -globulin (158 kDa), (D) chicken ovalbumin (44 kDa), (E) horse myoglobin (17 kDa) and (F) vitamin B₁₂ (1.35 kDa).

Table A1.3 Molecular weight standards (BioRad) used to predict molecular weights of FL-CHFR, Δ FHA1, Δ FHA2, Δ FHA3, FHA domain and C-terminal cysteine rich domain proteins by size exclusion chromatography.

	Molecular weight (kDa)	Elution volume (ml)	A280 (mAu)
Thyroglobulin (bovine)	670.000	51.35	39.89
γ -globulin (bovine)	158.000	67.34	45.21
Ovalbumin (chicken)	44.000	82.70	20.32
Myoglobin (horse)	17.000	93.70	39.32
Vitamin B12	1.350	110.94	50.19

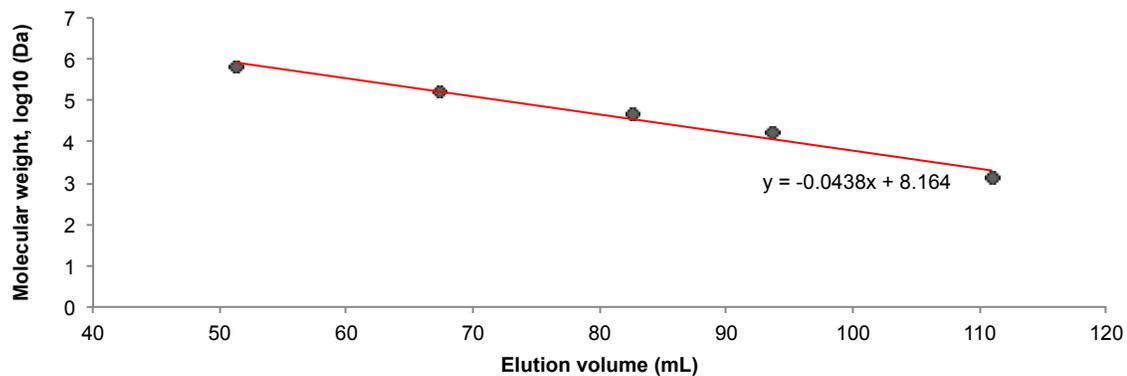


Figure A1.14 Calibration curve using molecular weight standards (BioRad) within a Highload 16/60 Superdex 200 Prep Grade column (GE Healthcare) to predict protein molecular weights. Plotted elution volumes and respective molecular weights (Table A1.3) are shown by the black diamonds, with a linear trend line (as shown in red) fitted to data points with the equation $y = -0.0438x + 8.164$.

A1.6 SEC-SAXS

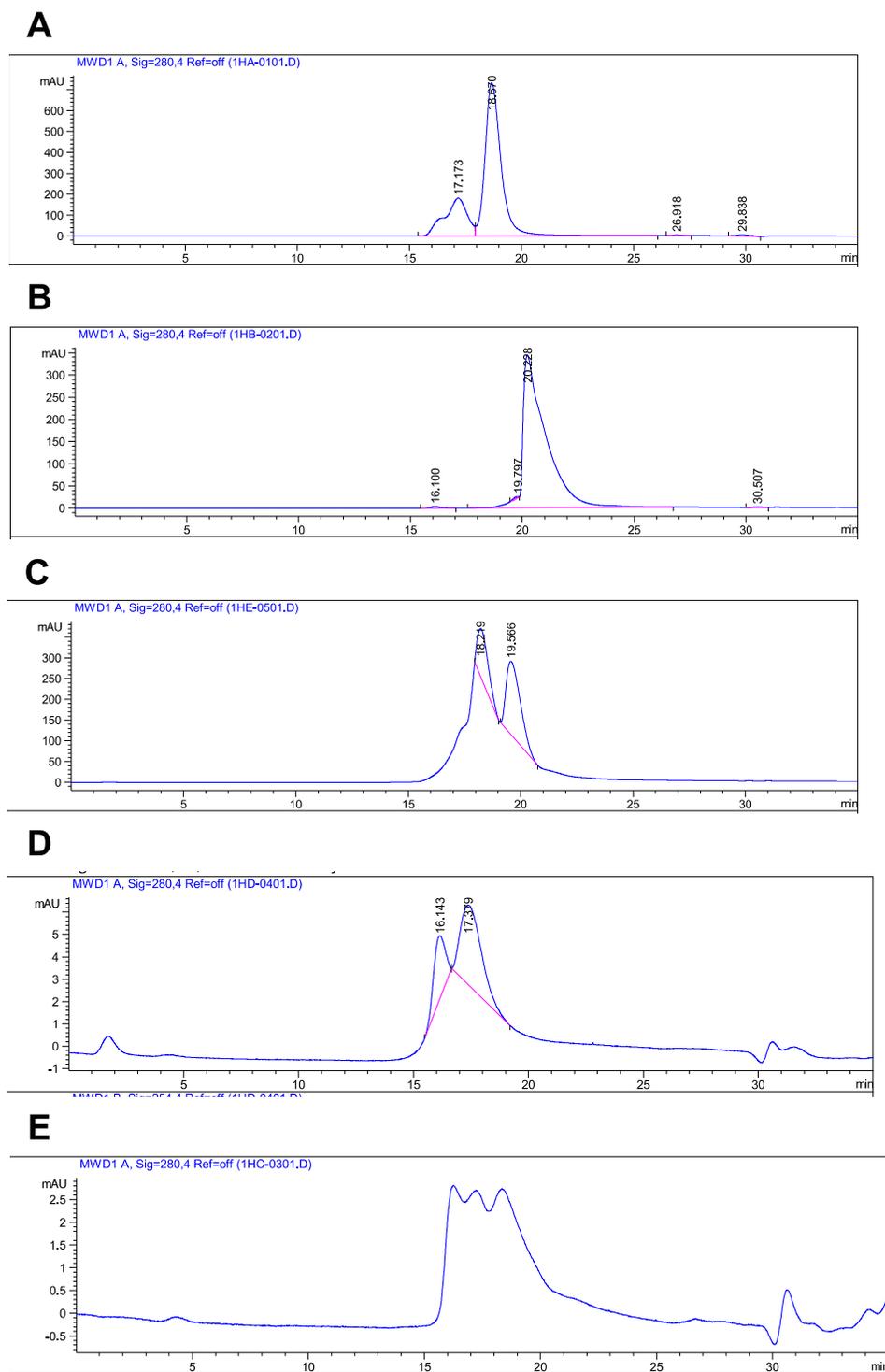


Figure A.1.15 Size Exclusion Chromatography (SEC) in line with SAXS (SEC-SAXS) performed using (A) BSA, (B) CHFR FHA domain, (C) CHFR C-terminus, (D) FL-CHFR and (E) Δ FHA-2 using a 4.6 mL KW-402.5 analytical column (Shodex). Retention times (minutes) are shown by numbers assigned to individual peaks (sectioned via magenta lines).

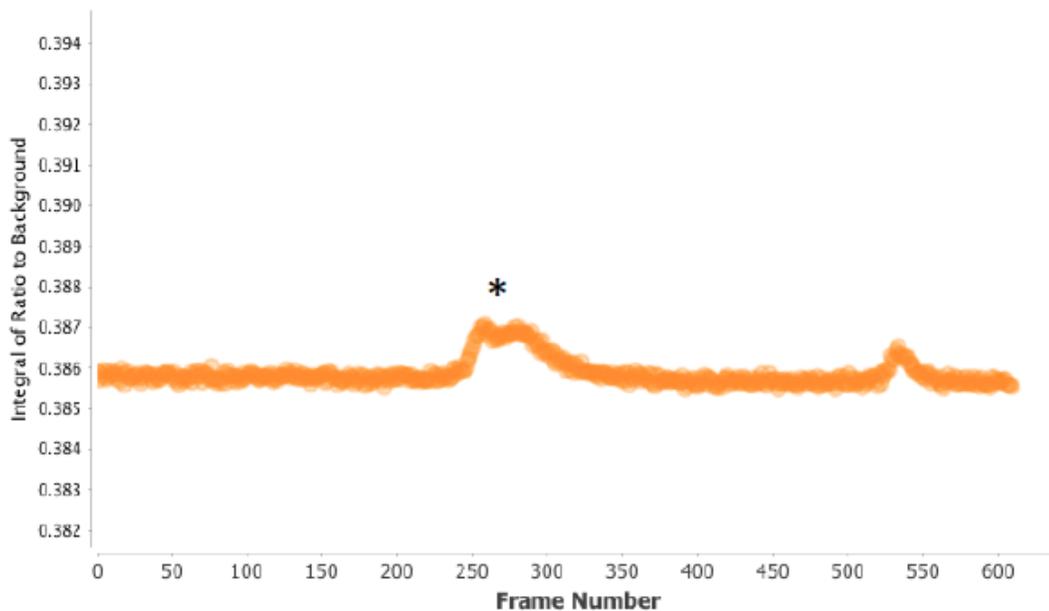
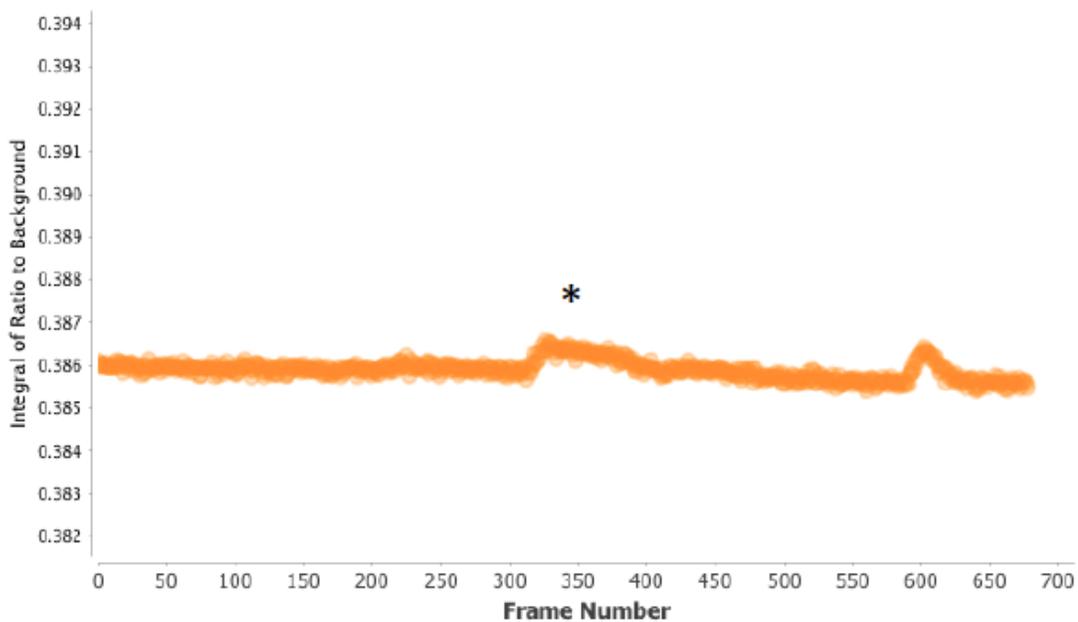
A**B**

Figure A1.16 Unsubtracted, integrated intensities from SEC-SAXS experiments performed using (A) FL-CHFR and (B) Δ FHA-2 proteins (resolved using a KW-403 analytical column), analysed using SCATTER (Rambo, 2017). Peaks corresponding to eluted FL-CHFR and Δ FHA-2 proteins are indicated by the asterisks (*).

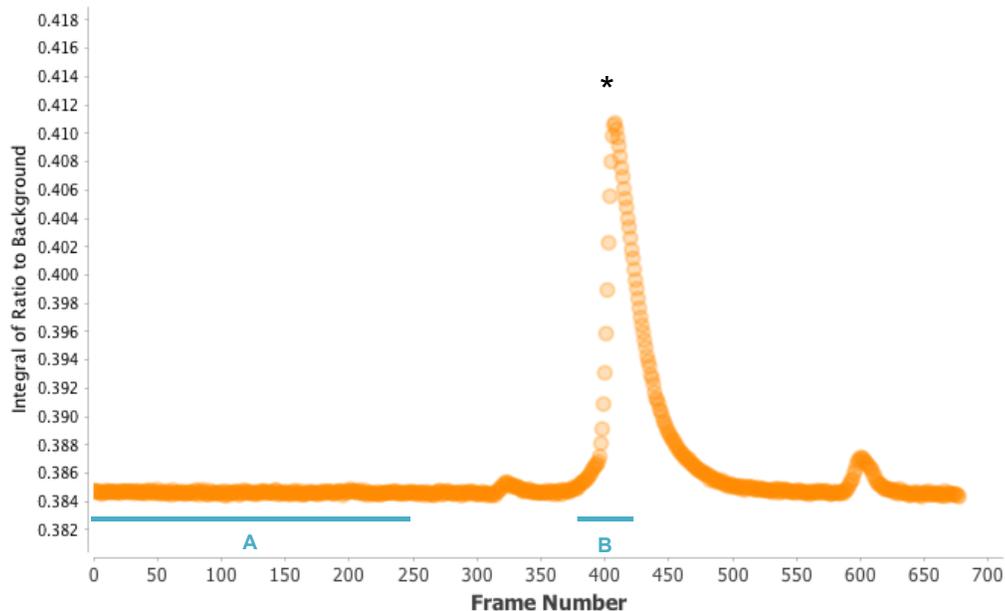


Figure A1.17 Unsubtracted, integrated intensities from SEC-SAXS experiment performed using the CHFR FHA domain (resolved using a KW-403 analytical column), analysed using SCATTER (Rambo, 2017). The peak corresponding to eluted FHA domain is indicated by the asterisks (*). The average of measured buffer intensities (A; frames 1 to 250) were subtracted from the average of eluted protein intensities (B; frames 380 to 430).

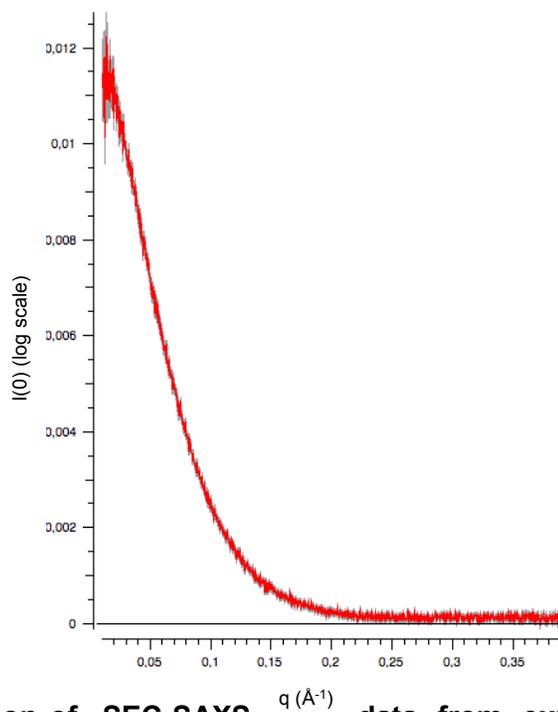


Figure A1.18 Evaluation of SEC-SAXS data from experiments performed using the CHFR FHA domain. The buffer subtracted profile of the FHA domain was plotted as the log of scattering intensity ($I(0)$) against magnitude of scattering intensity (or momentum transfer, q ; on an absolute scale). The well folded CHFR FHA domain (positive) scattering intensity approaches zero at higher q values; indicative of correct buffer matching which would otherwise be attributed to artificially high (positive) or negative $\log I(0)$ at higher q values (Skou *et al.*, 2014).

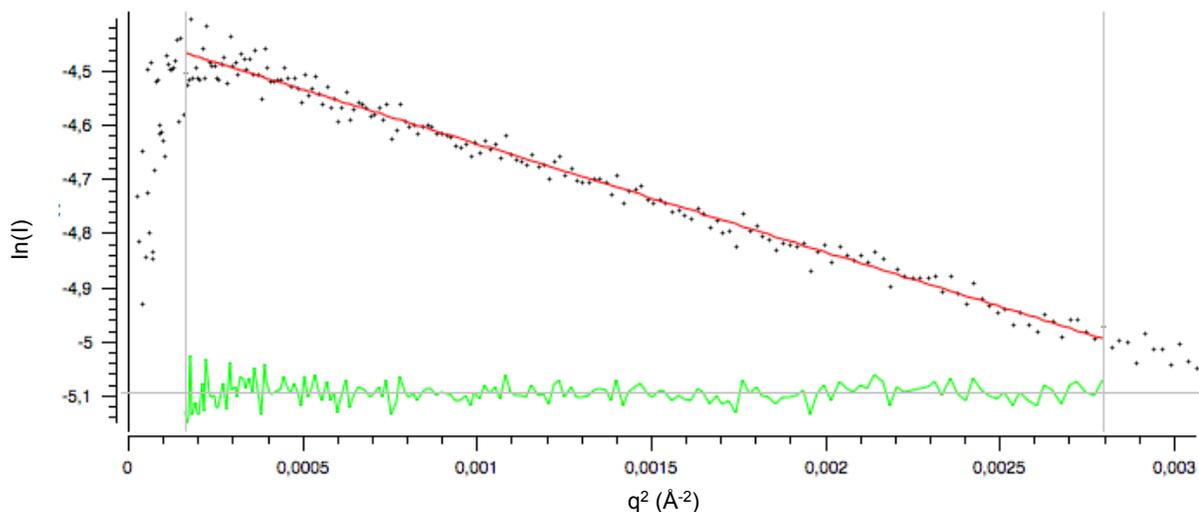


Figure A1.19 Guinier analysis of SEC-SAXS data from experiments performed using the CHFR FHA domain. The natural log of scattering intensities ($\ln(I)$) was plotted against the momentum transfer squared (q^2). The low- q values within the Guinier plot (black data points) are used to fit a red line; with residuals (corresponding to the fitted line's linearity) are also shown below (Skou *et al.*, 2014). Within the low q region, a lack of downturn is indicative of non-aggregated protein within solution (Skou *et al.*, 2014; Putnam *et al.*, 2007). From the Guinier analysis, calculated $I(0)$ and R_g values corresponding to the CHFR FHA domain are $0.012 \pm 3.3e-5$ (arbitrary units) and 24.51 ± 0.11 Å, respectively. A small number of noisy data points in the beginning of the plot have been omitted from the fit.

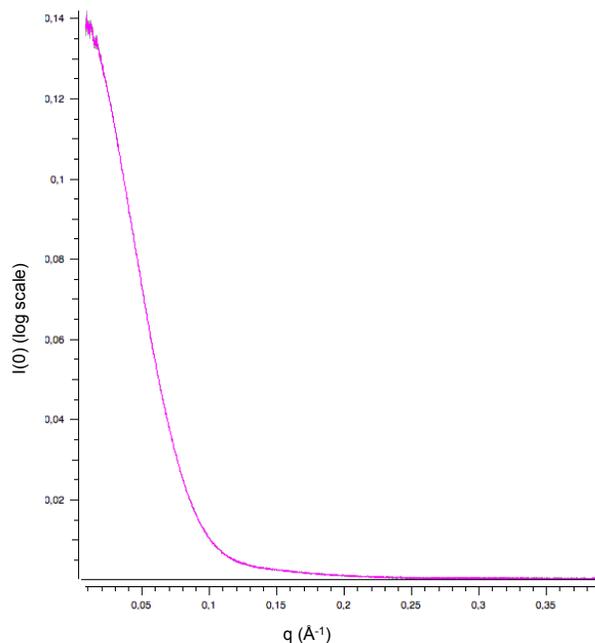


Figure A1.20 Evaluation of SEC-SAXS data from experiments performed using the BSA control domain. The buffer subtracted profile of the BSA protein control was plotted as the log of scattering intensity ($I(0)$) against magnitude of scattering intensity (or momentum transfer, q ; on an absolute scale). The well folded BSA (positive) scattering intensity approaches zero at higher q values; indicative of correct buffer matching which would otherwise be attributed to artificially high (positive) or negative log $I(0)$ at higher q values (Skou *et al.*, 2014).

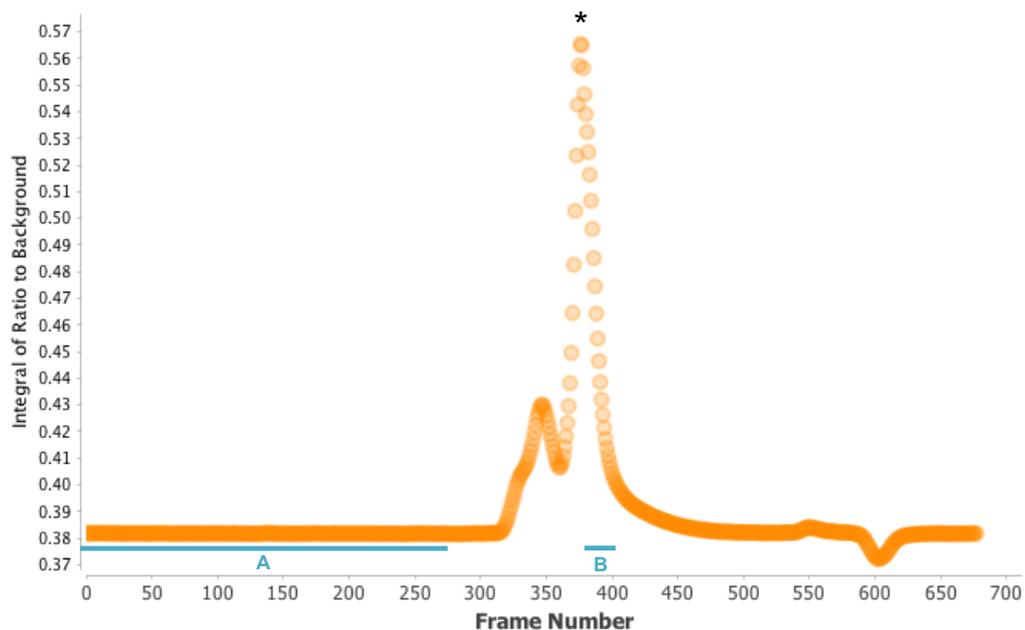


Figure A1.21 Unsubtracted, integrated intensities from SEC-SAXS experiment performed using a BSA control (resolved using a KW-403 analytical column), analysed using SCATTER (Rambo, 2017). The peak corresponding to eluted BSA protein is indicated by the asterisks (*). The average of measured buffer intensities (A; frames 1 to 280) were subtracted from the average of eluted protein intensities (B; frames 390 to 420).

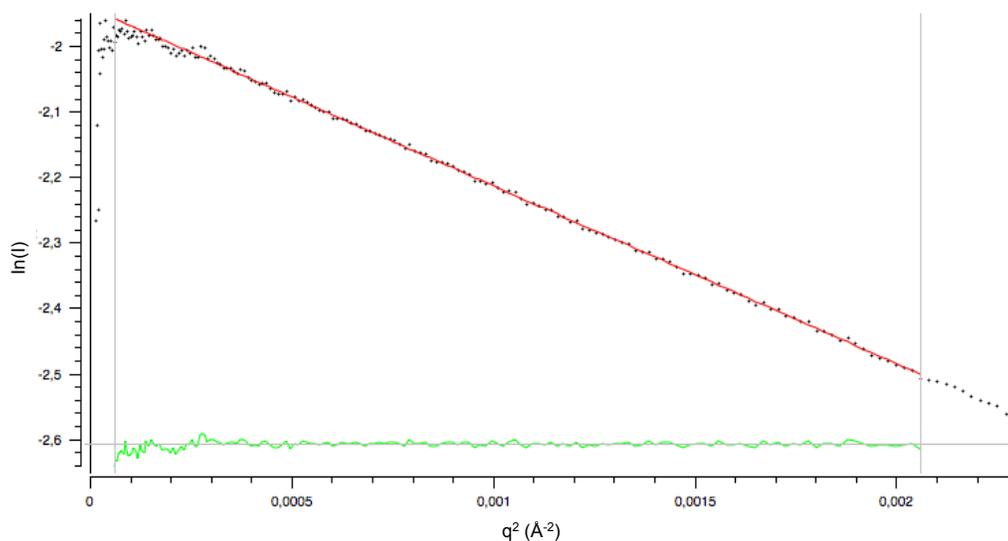


Figure A1.22 Guinier analysis of SEC-SAXS data from experiments performed using the BSA control. The natural log of scattering intensities ($\ln(I)$) was plotted against the momentum transfer squared (q^2). The low- q values within the Guinier plot (black data points) are used to fit a red line; with residuals (corresponding to the fitted line's linearity) are also shown below (Skou *et al.*, 2014). Within the low q region, a lack of downturn is indicative of non-aggregated protein within solution (Skou *et al.*, 2014; Putnam *et al.*, 2007). From the Guinier analysis, calculated $I(0)$ and R_g values corresponding to the BSA control are $0.14 \pm 7.5e-5$ (arbitrary units) and 28.52 ± 0.02 Å, respectively. A small number of noisy data points in the beginning of the plot have been omitted from the fit.

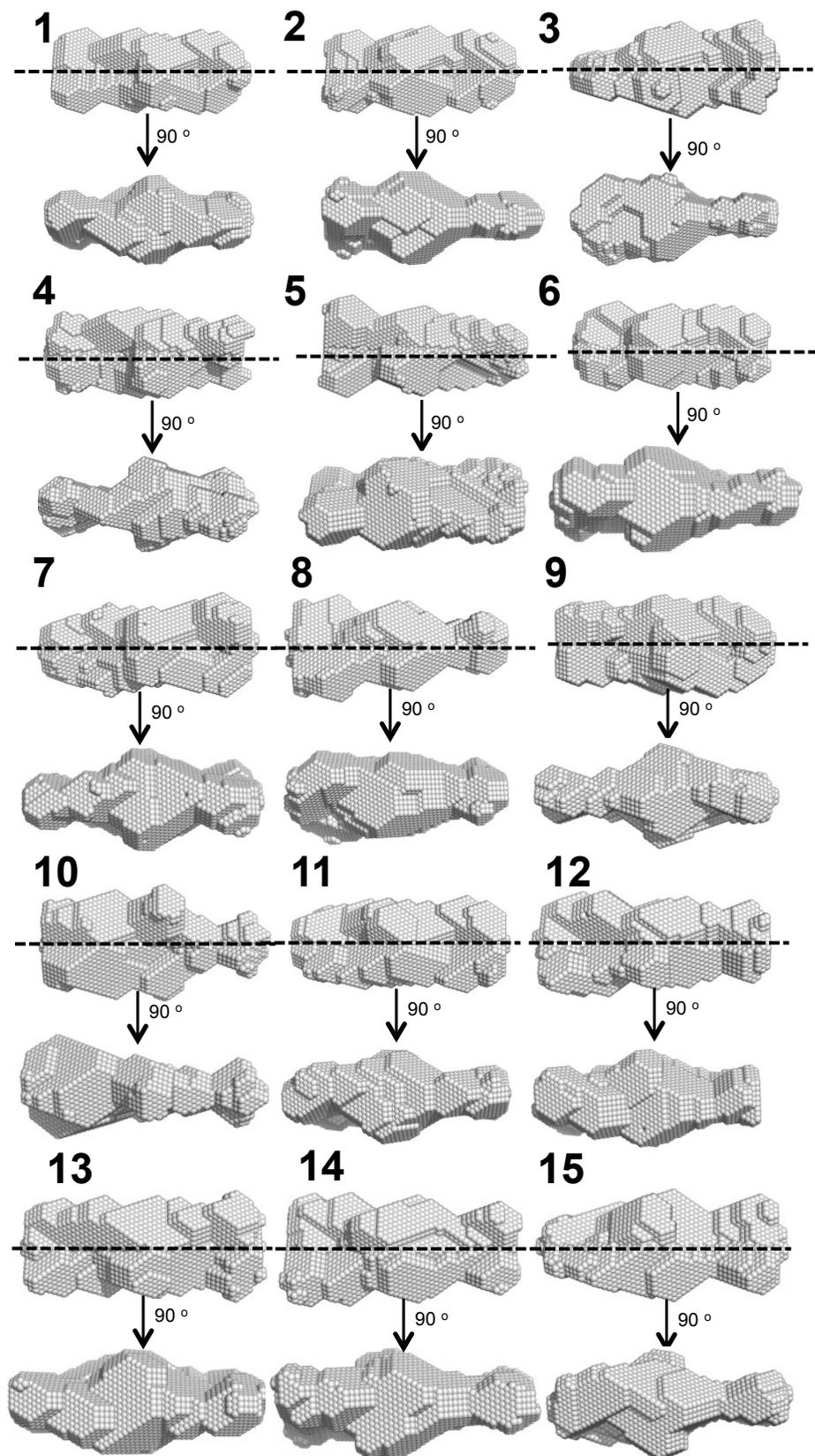


Figure A1.23 Dummy atom models of the solution structure of CHFR FHA domain dimer (13-180) generated using DAMMIF. Per model, the axis of P2 symmetry is indicated by the hashed lines. Individual models (1 to 20) are numbered, with individual dummy atoms, per particle, shown as grey spheres.

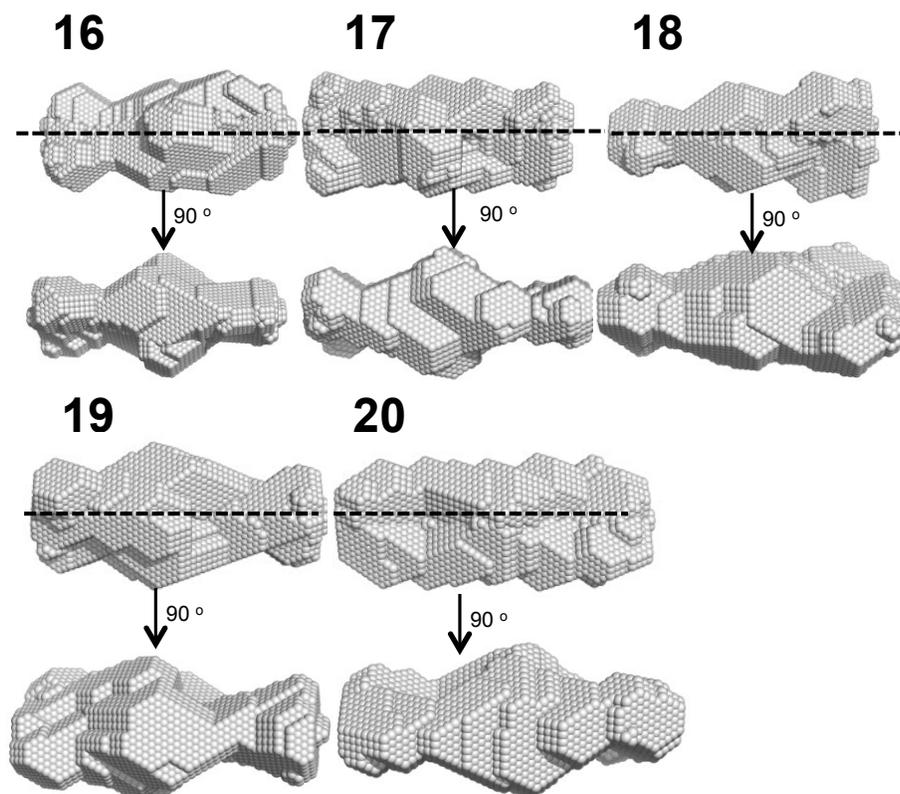


Figure A1.24 (...Continued). Dummy atom models of the solution structure of CHFR FHA domain dimer (13-180) generated using DAMMIF. Per model, the axis of P2 symmetry is indicated by the hashed lines. Individual models (1 to 20) are numbered, with individual dummy atoms, per particle, shown as grey spheres

A1.7 UBIQUITINATION ASSAYS

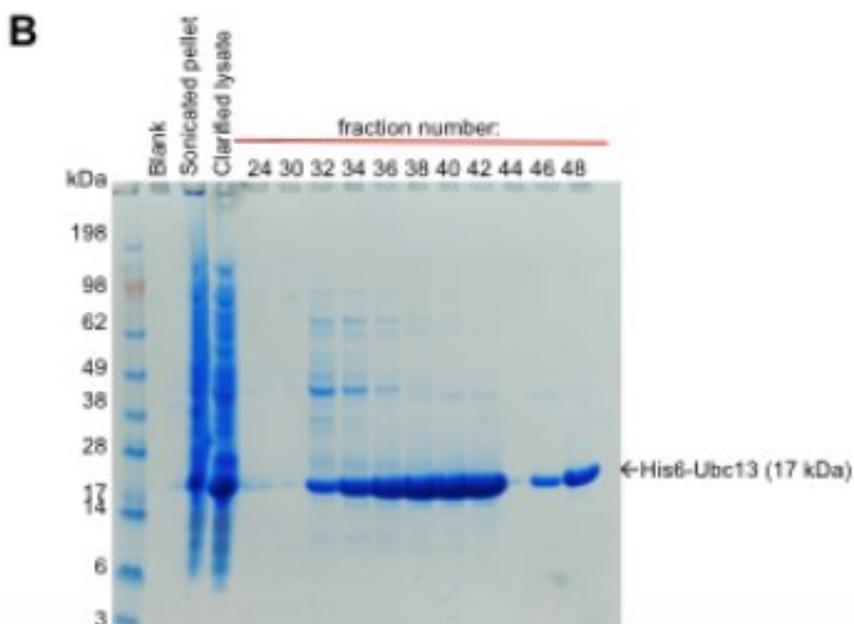
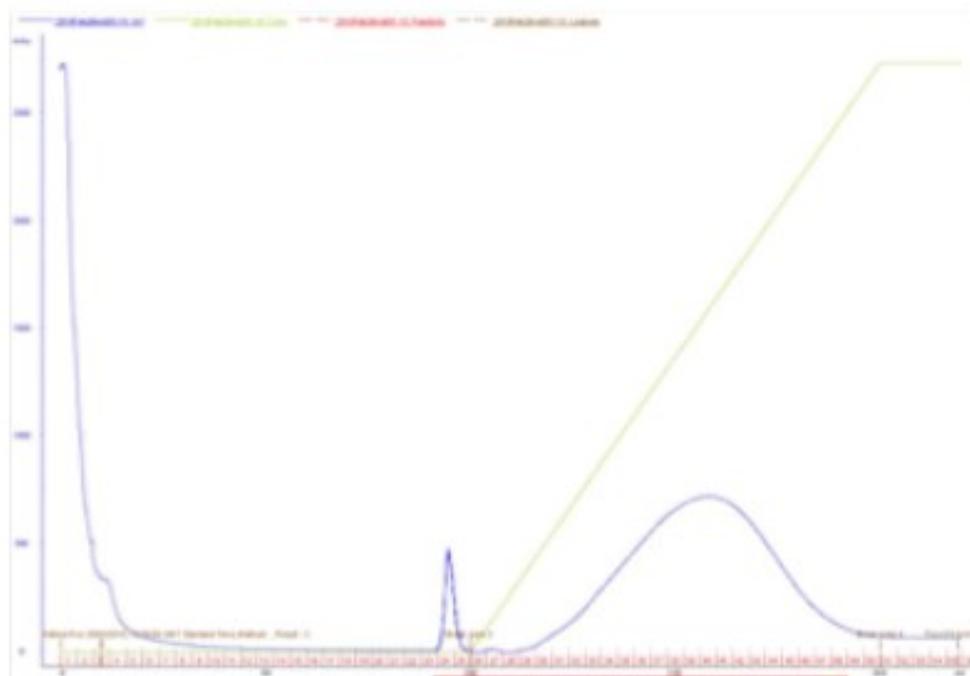


Figure A1.25 Purification of His-tagged Ubc13 ubiquitin conjugating enzyme by immobilized metal ion affinity chromatography. (A) Immobilized metal ion affinity chromatography was performed by loading clarified lysate onto a Histrap FF (Panel B, Lane 1), with the chromatogram indicating an elution peak (UV, A_{280} ; navy line); corresponding to a linear gradient of imidazole from 8 mM to 250 mM (concentration of high-imidazole buffer B; green line). (B) SDS-PAGE analysis of fractions taken from across the peak (see red, horizontal line in (A)) indicated His6-Ubc13 at 17 kDa.

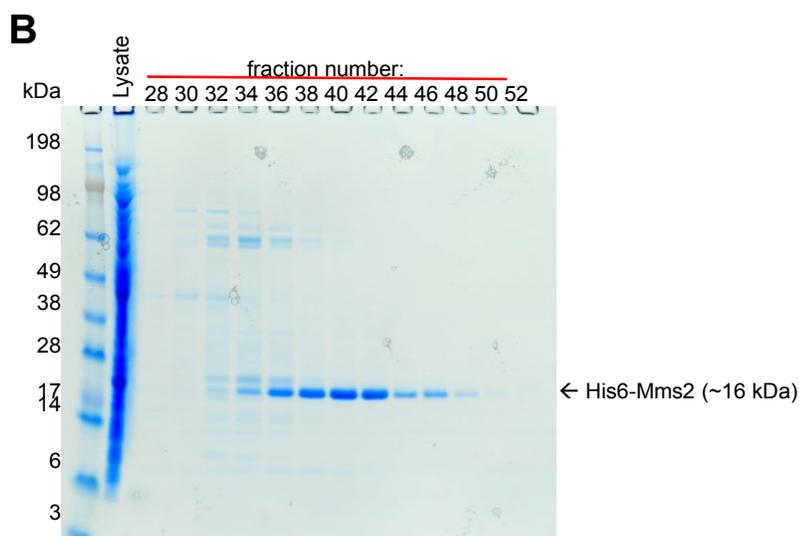
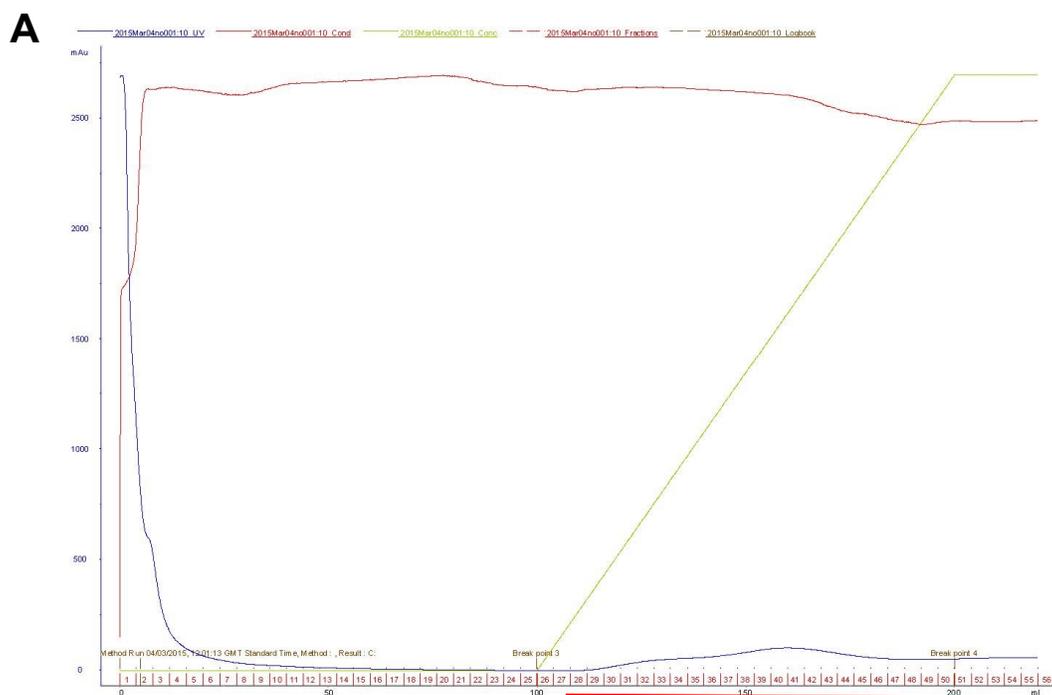


Figure A1.26 Purification of His-tagged Mms2 ubiquitin conjugating enzyme by immobilized metal ion affinity chromatography. (A) Immobilized metal ion affinity chromatography was performed by loading clarified lysate onto a Histrap FF (Panel B, Lane 1), with the chromatogram indicating an elution peak (UV, A_{280} ; navy line); corresponding to a linear gradient of imidazole from 8 mM to 250 mM (concentration of high-imidazole buffer B; green line). (B) SDS-PAGE analysis of fractions taken from across the peak (see red, horizontal line in (A)) indicated His6-Mms2 at 16 kDa.

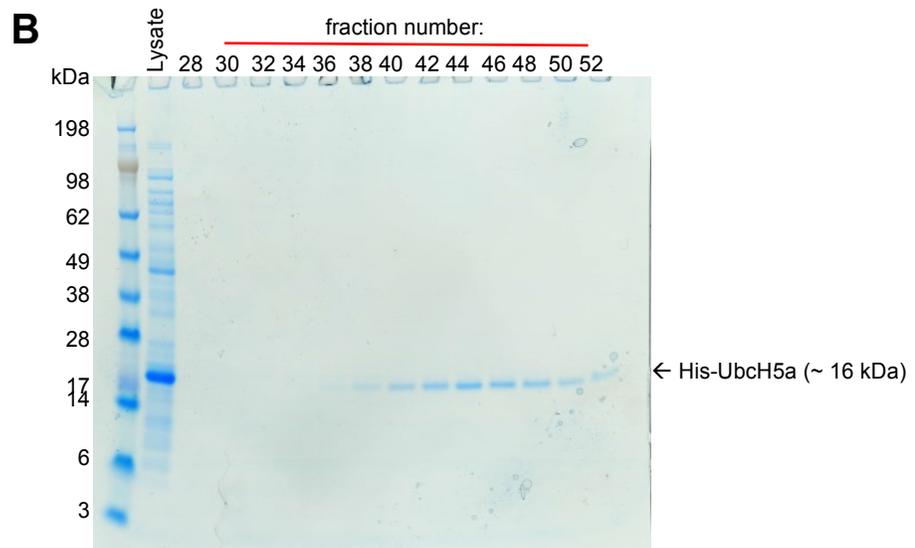
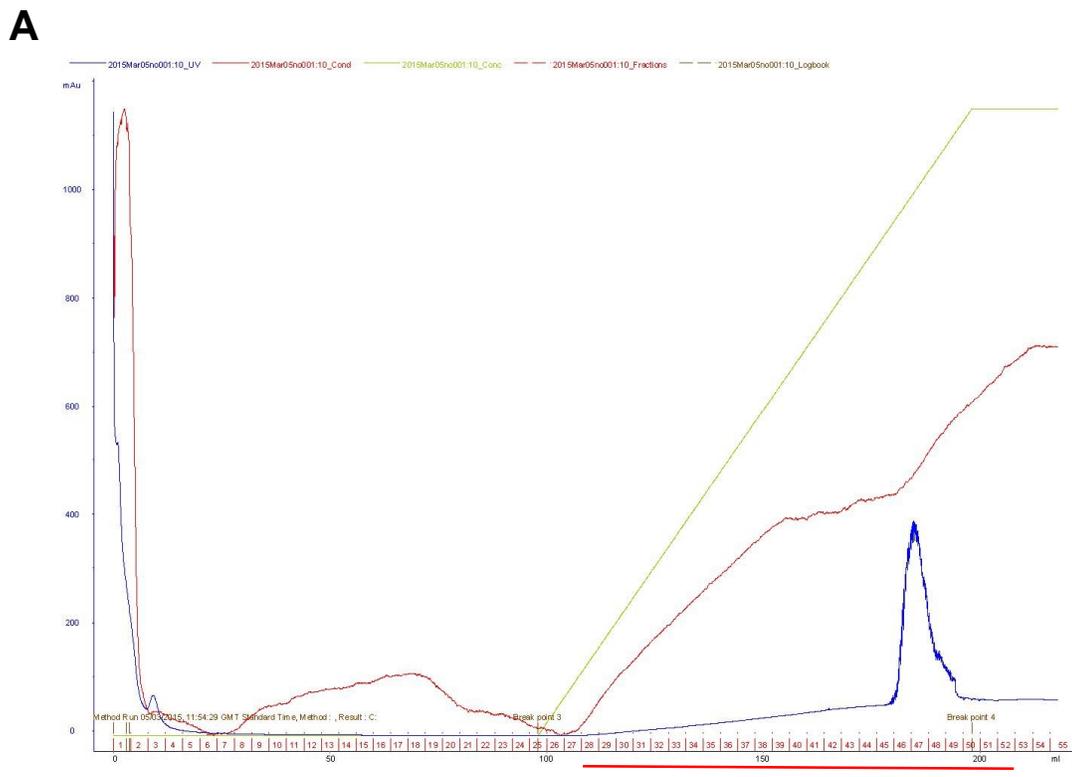


Figure A1.27 Purification of His-tagged UbcH5a ubiquitin conjugating enzyme by immobilized metal ion affinity chromatography. (A) Immobilized metal ion affinity chromatography was performed by loading clarified lysate onto a Histrap FF (Panel B, Lane 1), with the chromatogram indicating an elution peak (UV, A_{280} ; navy line); corresponding to a linear gradient of imidazole from 8 mM to 250 mM (concentration of high-imidazole buffer B; green line). (B) SDS-PAGE analysis of fractions taken from across the peak (see red, horizontal line in (A)) indicated His6-UbcH5a at 16 kDa.

Table A1.4 E2Scan Kit (Version 2) E2 ubiquitin conjugating enzymes (Ubiquigent) (Adapted from Anon, 2013).

E2 ubiquitin conjugating enzyme	Alternate name	Tag
HR6A	Ube2A	None
HR6B	Ube2B	None
UbcH10	Ube2C	T7
UbcH5a	Ube2D1	T7
UbcH5b	Ube2D2	T7
UbcH5c	Ube2D3	None
UbcH5d	Ube2D4	T7
UbcH6	Ube2E1	None
UbcH8	Ube2E2	T7
UbcH9	Ube2E3	None
NCE2	Ube2F	T7
Ubc7	Ube2G1	T7
UbcH2	Ube2G2	None
UbcH2	Ube2H	None
Ubc9	Ube2I	None
NCUBE1	6His-Ube2J1	His-T7
NCUBE2	Ube2J2	T7
Ubc1	Ube2K	None
UbcH7	Ube2L3	None
UbcH8	Ube2L6	None
Ubc12	Ube2M	None
Ubc13	Ube2N	None
Ubc13/ Uev1A	Ube2N/ Ube2V1	None/ T7
Ubc13/ Mms2	Ube2N/ Ube2V2	None/ None
NICE-5	His6-Ube2Q	His-T7
Ube2Q2	None	None

Table A1.4 E2Scan Kit (Version 2) E2 ubiquitin conjugating enzymes (...Continued) (Ubiquigent) (Adapted from Anon, 2013).

E2 ubiquitin conjugating enzyme	Alternate name	Tag
CDC34	Ube2R1	T7
CDC34B	Ube2R2	T7
E2-EPF	Ube2S	T7
HSPC-150	Ube2T	None
Uev1A	Ube2V1	T7
Mms2	Ube2V2	None
Ubc16	His6-Ube2W	His-T7
USE1	His6-Ube2Z	His-T7

Appendix 2

I - Examples: homology modelling scripts and files

A2.1 Basic modeling alignment file: RING_Zn.ali

```
>P1;CHFR_RING
sequence:CHFR_RING:1:A:52:A:CHFR:human: :
NELQCIICSEYFIEAVTLN-CAHSFCSYICINEWMKRKIECPICRKDI..*
```

```
>P1;4ayc
structureX:4ayc:399:A:46:A:RING: : :
ETLTICIICQDLLHDCVSLQPCMHFTFCAACYSGWMERSLCPTRCPV..*
```

A2.2 Basic (single template) modelling python script: CHFRring_4ayc.py

```
from modeller import *
from modeller.automodel import *
log.verbose()
env = environ()
env.io.atom_files_directory = ['.', '../atom_files']
env.io.hetatm = True
a = automodel(env,
              alnfile = 'RING_Zn.ali', # alignment filename
              knowns = '4ayc',        # codes of the templates
              sequence = 'CHFR_RING',  # code of the target
              assess_methods=(assess.DOPE,
                              #soap_protein_od.Scorer(),
                              assess.GA341))
a.starting_model= 1
a.ending_model = 100 # Index of last model (1-100 models produced)
a.make()
```

A2.3 Multi template modelling python script – (a) *salign*: *salign.py*

```
# Homology modeling with ligand transfer from the template
from modeller import *          # Load standard Modeller classes
from modeller.automodel import * # Load the automodel class
log.verbose() # request verbose output
env = environ() # create a new MODELLER environment to build this model in
# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']
# Read in HETATM records from template PDBs
env.io.hetatm = True
aln = alignment(env)
for (code, chain) in (('4ayc', 'A'), ('4s3o', 'B'), ('5d0i', 'A'), ('4qpl', 'A'),
('3ng2', 'A'), ('3zni', 'A')):
    mdl = model(env, file=code, model_segment=('FIRST:'+chain, 'LAST:'+chain))
    aln.append_model(mdl, atom_files=code, align_codes=code+chain)
for (weights, write_fit, whole) in (((1., 0., 0., 0., 1., 0.), False, True),
((1., 0.5, 1., 1., 1., 0.), False, True),
((1., 1., 1., 1., 1., 0.), True, False)):
    aln.salign(rms_cutoff=3.5, normalize_pp_scores=False,
rr_file='$ (LIB)/asl.sim.mat', overhang=30,
gap_penalties_1d=(-450, -50),
gap_penalties_3d=(0, 3), gap_gap_score=0, gap_residue_score=0,
dendrogram_file='fm00495.tree',
alignment_type='tree', # If 'progressive', the tree is not
# computed and all structures will be
# aligned sequentially to the first
feature_weights=weights, # For a multiple sequence alignment only
# the first feature needs to be non-zero
improve_alignment=True, fit=True, write_fit=write_fit,
write_whole_pdb=whole, output='ALIGNMENT QUALITY')
aln.write(file='fm00495.pap', alignment_format='PAP')
aln.write(file='fm00495.ali', alignment_format='PIR')
aln.salign(rms_cutoff=1.0, normalize_pp_scores=False,
rr_file='$ (LIB)/asl.sim.mat', overhang=30,
gap_penalties_1d=(-450, -50), gap_penalties_3d=(0, 3),
gap_gap_score=0, gap_residue_score=0, dendrogram_file='lis3A.tree',
alignment_type='progressive', feature_weights=[0]*6,
improve_alignment=False, fit=False, write_fit=True,
write_whole_pdb=False, output='QUALITY')
```

A2.4 Multi template modelling python script – (b) align.block: align_block.py

```
from modeller import *
log.verbose()
env = environ()
# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']
# Read in HETATM records from template PDBs
env.io.hetatm = True
env.libs.topology.read(file='$ (LIB)/top_heav.lib')
# Read aligned structure(s):
aln = alignment(env)
aln.append(file='fm00495.ali', align_codes='all')
aln_block = len(aln)
# Read aligned sequence(s):
aln.append(file='RING_Zn.ali', align_codes='RING_Zn')
# Structure sensitive variable gap penalty sequence-sequence alignment:
aln.salign(output='', max_gap_length=20,
           gap_function=True, # to use structure-dependent gap penalty
           alignment_type='PAIRWISE', align_block=aln_block,
           feature_weights=(1., 0., 0., 0., 0., 0.), overhang=0,
           gap_penalties_1d=(-450, 0),
           gap_penalties_2d=(0.35, 1.2, 0.9, 1.2, 0.6, 8.6, 1.2, 0., 0.),
           similarity_flag=True)
aln.write(file='RING-mult.ali', alignment_format='PIR')
aln.write(file='RING-mult.pap', alignment_format='PAP')
```

A2.5 Multi template modelling alignment file – (b) align.block: RING_Zn.ali

```
>P1;RING_Zn
sequence:RING_Zn:1:A:49:A:CHFR:human: :
ETLTTCIICQDLLHDCVSLQ-----PCMHTFCAACYSGWMERSSL-CPTCRCPV...*

>P1;4ayc
structureX:4ayc:399:A:446:A:RING: : :
NELQCIICSEYFIEAVTLN-----CAHSFCSYICINEWMKRKI-ECPICRKDI...*

>P1;5d0i
sequence:5d0i:371:A:341:A:CHFR:human: :
TDEKCTICLSMLEDGEDVRRL----PCMHLFHQLCVDQWLAMSKK-CPICRVDI...*

>P1;4QPL
sequence:4QPL:32:A:79:A:CHFR:human: :
TVPECAICLQTCVHPVSL-----PCKHVFCYLCVKGASWLGKR-CALCRQEI...*

>P1;4S3O
sequence:4S3O:47:B:96:B:CHFR:human: :
SELMCPICLDMLKNTMTTKE-----CLHRFCADCIITALRSGNKECPTCRKKL...*

>P1;3ZNI
sequence:3ZNI:369:A:417:A:CHFR:human: :
TFQLCKICAENDKDVKIE-----PCGHLMCTSLTAWQESDGQGPCFRCRCEI...*

>P1;3NG2
sequence:3ZNI:132:A:186:A:CHFR:human: :
GTVSCPICMDGYSEIVQNGRLIVSTE-CGHVFCQCLRDSLKNANT-CPTCRKKI...*
```

A2.6 Multi template modelling python script – (c) model_multi: model_multi.py

```
from modeller import *
from modeller.automodel import *
log.verbose()
env = environ()
# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']
# Read in HETATM records from template PDBs
env.io.hetatm = True
a = automodel(env, alnfile='RING-mult.ali',
              knowns=('4aycA', '4qplA', '4s3oB', '5d0iA', '3zniA', '3ng2A'),
              sequence='RING_Zn',
              assess_methods=(assess.DOPE,
                              #soap_protein_od.Scorer(),
                              assess.GA341))

a.starting_model = 1
a.ending_model = 100
a.make()
```

A2.7 Model (and template) evaluation python script: evaluate_template4ayc.py

```
PDB file (template): 4acy.pdb
from modeller import *
from modeller.scripts import complete_pdb
log.verbose()# request verbose output
env = environ()
env.libs.topology.read(file='$ (LIB)/top_heav.lib') # read topology
env.libs.parameters.read(file='$ (LIB)/par.lib') # read parameters
# read model file
mdl = complete_pdb(env, '4ayc')
# Assess with DOPE:
s = selection(mdl)# all atom selection
s.assess_dope(output='ENERGY_PROFILE NO_REPORT', file='CHFR_RING.profile',
             normalize_profile=True, smoothing_window=15)
```

A2.8 Loop refinement python script: Loop_refine.py

```
from modeller import *
from modeller.automodel import *
log.verbose()
env = environ()
env.io.atom_files_directory = './../atom_files'
env.io.hetatm = True
class MyLoop(loopmodel):
    return selection(self.residue_range('6', '16'))
m = MyLoop(env,
           inimodel='RING_Zn.B99990106', # initial model of the target
           sequence='RING_Zn', # code of the target
           assess_methods=(assess.DOPE,
                           #soap_protein_od.Scorer(),
                           assess.GA341))

m.loop.starting_model= 1
m.loop.ending_model = 500
m.loop.md_level = refine.slow # refine.slow: higher quality models,
m.make()
```

II - Zinc AMBER Force Field (ZAFF) topology and coordinate file production

A2.9 Commands within the -tsch terminal shell (N.B:- residue numbers are in reference to the CHFR RING model 10 and not the full length CHFR protein).

```
setenv AMBERHOME /Users/lfrnal/Desktop/Amber_tools_NEW/amber14/
/Users/lfrnal/Desktop/Amber_tools_NEW/amber14/bin/tleap -s -f

/Users/lfrnal/Desktop /Amber_tools_NEW/amber14/dat/leap/cmd/leaprc.ff14SB

source leaprc.ff14SB #Source the ff14SB force field
addAtomTypes { { "ZN" "Zn" "sp1" } { "S1" "S" "sp1" } { "N1" "N" "sp1" } }
loadoff atomic_ions.lib
loadamberparams frcmod.ions11sm_hfe_tip3p
loadamberprep ZAFF.prep
loadamberparams ZAFF.frcmod

mol = loadPdb RING_Zn106.pdb
bond mol.48.ZN mol.5.SG #bond zinc ion (1) with SG atom or residue 5 (Cys)
bond mol.48.ZN mol.8.SG #bond zinc ion (1) with SG atom or residue 8 (Cys)
bond mol.48.ZN mol.26.SG #bond zinc ion (1) with SG atom or residue 26 (Cys)
bond mol.48.ZN mol.29.SG #bond zinc ion (1) with SG atom or residue 29 (Cys)

bond mol.49.ZN mol.21.SG #bond zinc ion (2) with SG atom or residue 21 (Cys)
bond mol.49.ZN mol.23.NE2 #bond zinc ion (2) with SG atom or residue 23 (His)
bond mol.49.ZN mol.40.SG #bond zinc ion (2) with SG atom or residue 40 (Cys)
bond mol.49.ZN mol.43.SG #bond zinc ion (2) with SG atom or residue 43 (Cys)

savepdb mol RING_Zn081_dry.pdb #Save the pdb file; dry
saveamberparm mol RING_Zn106_GMX.prmtop RING_Zn106_GMX.inpcrd
quit
```

III - Molecular dynamics: input files

A2.10 Addition of ions: *ions.mdp*

```
; ions.mdp - used as input into grompp to generate ions.tpr

; Parameters describing what to do, when to stop and what to save
integrator      = steep          ; Algorithm (steep = steepest descent minimization)
emtol           = 1000.0        ; Stop minimization when the maximum force < 1000.0
kJ/mol/nm
emstep         = 0.01           ; Energy step size
nsteps         = 50000          ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to
calculate the interactions
nstlist        = 1             ; Frequency to update the neighbor list and long range
forces
ns_type        = grid          ; Method to determine neighbor list (simple, grid)
rlist          = 1.0           ; Cut-off for making neighbor list (short range forces)
coulombtype    = PME           ; Treatment of long range electrostatic interactions
rcoulomb       = 1.0           ; Short-range electrostatic cut-off
rvdw           = 1.0           ; Short-range Van der Waals cut-off
pbc            = xyz           ; Periodic Boundary Conditions (yes/no)
```

A2.11 Energy minimization: *minim.mdp*

```
; minim.mdp - used as input into grompp to generate em.tpr
; Parameters describing what to do, when to stop and what to save
integrator      = steep          ; Algorithm (steep = steepest descent minimization)
emtol           = 1000.0        ; Stop minimization when the maximum force < 1000.0
kJ/mol/nm
emstep         = 0.01           ; Energy step size
nsteps         = 50000          ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to
calculate the interactions
nstlist        = 1             ; Frequency to update the neighbor list and long range
forces
ns_type        = grid          ; Method to determine neighbor list (simple, grid)
rlist          = 1.0           ; Cut-off for making neighbor list (short range forces)
coulombtype    = PME           ; Treatment of long range electrostatic interactions
rcoulomb       = 1.0           ; Short-range electrostatic cut-off
rvdw           = 1.0           ; Short-range Van der Waals cut-off
pbc            = xyz           ; Periodic Boundary Conditions (yes/no)
```

A2.12 NVT ensemble: nvt.mdp

```
define          = -DPOSRES      ; position restrain the protein

; Run parameters
integrator      = md            ; leap-frog integrator
nsteps          = 250000        ; 2 * 250000 = 500 ps
dt              = 0.002         ; 2 fs

; Output control
nstxout         = 100           ; save coordinates every 0.2 ps
nstvout         = 100           ; save velocities every 0.2 ps
nstenergy       = 100           ; save energies every 0.2 ps
nstlog          = 100           ; update log file every 0.2 ps

; Bond parameters
continuation    = no            ; first dynamics run
constraint_algorithm = lincs     ; holonomic constraints
constraints     = all-bonds     ; all bonds (even heavy atom-H bonds) constrained
lincs_iter      = 1             ; accuracy of LINCS
lincs_order     = 4             ; also related to accuracy

; Neighborsearching
ns_type         = grid          ; search neighboring grid cells
nstlist         = 5             ; 10 fs
rlist           = 1.0           ; short-range neighborlist cutoff (in nm)
rcoulomb        = 1.0           ; short-range electrostatic cutoff (in nm)
rvdw            = 1.0           ; short-range van der Waals cutoff (in nm)

; Electrostatics
coulombtype     = PME           ; Particle Mesh Ewald for long-range electrostatics
pme_order       = 4             ; cubic interpolation
fourierspacing = 0.16          ; grid spacing for FFT

; Temperature coupling is on
tcoupl         = V-rescale      ; modified Berendsen thermostat
tc-grps        = Protein Non-Protein ; two coupling groups - more
accurate
tau_t          = 0.1 0.1        ; time constant, in ps
ref_t          = 300 300        ; reference temperature, one for each group, in K

; Pressure coupling is off
pcoupl         = no            ; no pressure coupling in NVT

; Periodic boundary conditions
pbc            = xyz           ; 3-D PBC

; Dispersion correction
DispCorr       = EnerPres      ; account for cut-off vdW scheme

; Velocity generation
gen_vel        = yes           ; assign velocities from Maxwell distribution
gen_temp       = 300           ; temperature for Maxwell distribution
gen_seed       = -1            ; generate a random seed
```

A2.13 NPT ensemble: npt.mdp

```
define          = -DPOSRES      ; position restrain the protein

; Run parameters
integrator      = md            ; leap-frog integrator
nsteps         = 250000        ; 2 * 250000 = 500 ps
dt             = 0.002         ; 2 fs

; Output control
nstxout        = 100          ; save coordinates every 0.2 ps
nstvout        = 100          ; save velocities every 0.2 ps
nstenergy      = 100          ; save energies every 0.2 ps
nstlog         = 100          ; update log file every 0.2 ps

; Bond parameters
continuation   = yes          ; Restarting after NVT
constraint_algorithm = lincs    ; holonomic constraints
constraints    = all-bonds     ; all bonds (even heavy atom-H bonds) constrained
lincs_iter     = 1            ; accuracy of LINCS
lincs_order    = 4            ; also related to accuracy

; Neighborsearching
ns_type        = grid         ; search neighboring grid cells
nstlist        = 5            ; 10 fs
rlist          = 1.0          ; short-range neighborlist cutoff (in nm)
rcoulomb       = 1.0          ; short-range electrostatic cutoff (in nm)
rvdw           = 1.0          ; short-range van der Waals cutoff (in nm)

; Electrostatics
coulombtype    = PME          ; Particle Mesh Ewald for long-range electrostatics
pme_order      = 4            ; cubic interpolation
fourierspacing = 0.16        ; grid spacing for FFT

; Temperature coupling is on
tcoupl        = V-rescale     ; modified Berendsen thermostat
tc-grps       = Protein Non-Protein ; two coupling groups - more
accurate
tau_t         = 0.1 0.1       ; time constant, in ps
ref_t         = 300 300       ; reference temperature, one for each group, in K

; Pressure coupling is on
pcoupl        = Parrinello-Rahman ; Pressure coupling on in NPT
pcoupltype    = isotropic      ; uniform scaling of box vectors
tau_p         = 2.0           ; time constant, in ps
ref_p         = 1.0           ; reference pressure, in bar
compressibility = 4.5e-5       ; isothermal compressibility of water, bar^-1
refcoord_scaling = com

; Periodic boundary conditions
pbc           = xyz           ; 3-D PBC

; Dispersion correction
DispCorr      = EnerPres      ; account for cut-off vdW scheme

; Velocity generation
gen_vel       = no           ; Velocity generation is off
```

A2.14 Production MD: md.mdp

```
; Run parameters
integrator      = md          ; leap-frog integrator
nsteps         = 35000000    ;
dt             = 0.002       ; 2 fs

; Output control
nstxout        = 5000        ; save coordinates every 10.0 ps
nstvout        = 5000        ; save velocities every 10.0 ps
nstenergy      = 5000        ; save energies every 10.0 ps
nstlog         = 5000        ; update log file every 10.0 ps
nstxout-compressed = 5000    ; save compressed coordinates every 10.0
ps

; nstxout-compressed replaces nstxtcout
compressed-x-grps = System    ; replaces xtc-grps

; Bond parameters
continuation    = yes        ; Restarting after NPT
constraint_algorithm = lincs    ; holonomic constraints
constraints     = all-bonds    ; all bonds (even heavy atom-H bonds)
constrained
lincs_iter      = 1          ; accuracy of LINCS
lincs_order     = 4          ; also related to accuracy

; Neighborsearching
cutoff-scheme   = Verlet
ns_type         = grid        ; search neighboring grid cells
nstlist        = 10          ; 20 fs, largely irrelevant with Verlet
scheme
rcoulomb       = 1.0         ; short-range electrostatic cutoff (in nm)
rvdw           = 1.0         ; short-range van der Waals cutoff (in nm)

; Electrostatics
coulombtype    = PME          ; Particle Mesh Ewald for long-range
electrostatics
pme_order      = 4            ; cubic interpolation
fourierspacing = 0.16        ; grid spacing for FFT

; Temperature coupling is on
tcoupl        = V-rescale     ; modified Berendsen thermostat
tc-grps       = Protein Non-Protein ; two coupling groups - more
accurate
tau_t         = 0.1 0.1       ; time constant, in ps
ref_t         = 300 300       ; reference temperature, one for each group, in
K

; Pressure coupling is on
pcoupl        = Parrinello-Rahman ; Pressure coupling on in NPT
pcoupltype    = isotropic       ; uniform scaling of box vectors
tau_p         = 2.0            ; time constant, in ps
ref_p         = 1.0            ; reference pressure, in bar
compressibility = 4.5e-5        ; isothermal compressibility of
water, bar^-1
; Periodic boundary conditions
pbc           = xyz           ; 3-D PBC

; Dispersion correction
DispCorr      = EnerPres      ; account for cut-off vdW scheme
gen_vel       = no            ; Velocity generation is off
```

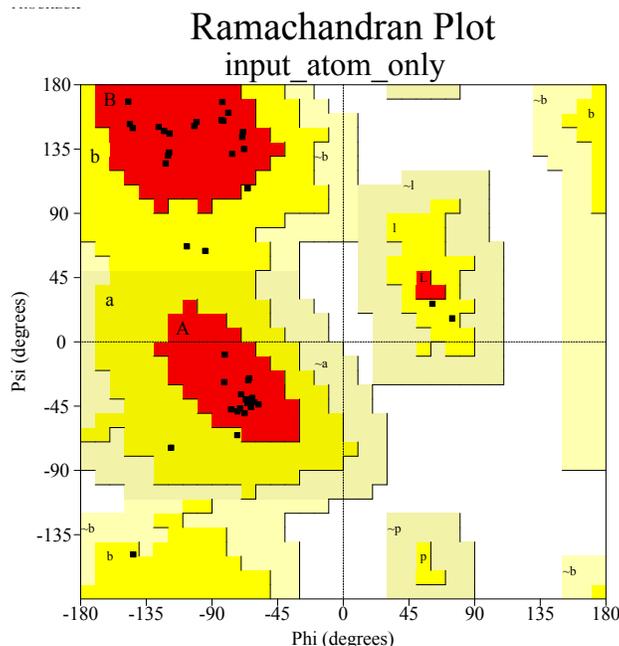


Figure A2.1 Ramachandran plot produced for the CHFR-RING model (106). PROCHECK statistical analysis indicates 80.5% of residues are within the most favoured regions (A, B and L); with 19.5% of residues within additionally allowed regions (a, b, l, p). No residues are within the generously allowed or disallowed regions.

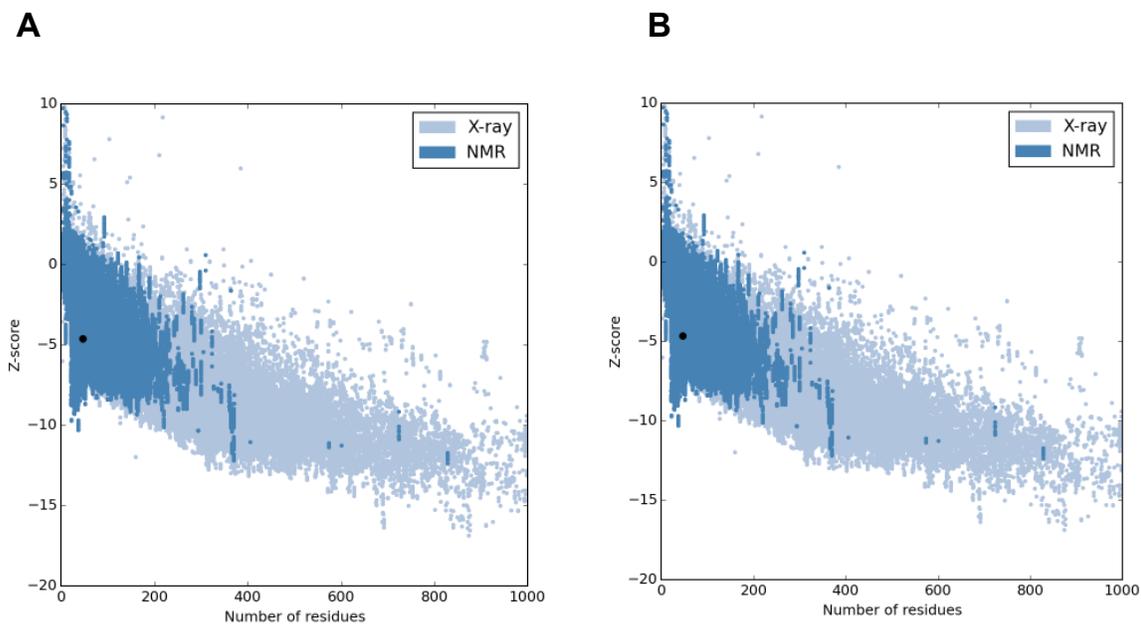


Figure A2.2 Graphical outputs from (A) CHFR RING model 106 and (B) template (PDB: 4AYC) evaluation using ProSa. Z-scores corresponding to NMR derived datasets are depicted by the dark blue colouration, whilst the X-ray crystallographic structural counterparts are shown in light blue. The black dots correspond to Z-score positions of (A) the RING CHFR model and (B) 4AYC PDB template, with respect to known high-quality datasets previously described.

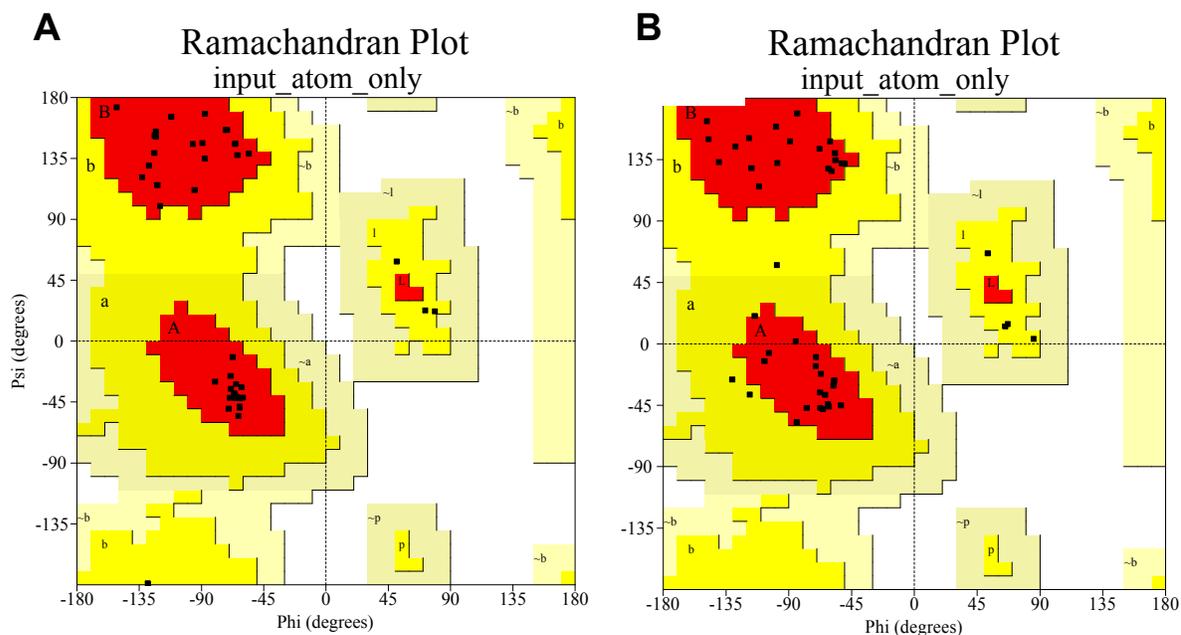


Figure A2.3 Ramachandran plot produced for the (A) RNF8 model (641), in comparison to the deposited structure (PDB: 4AYC). (A) PROCHECK statistical analysis indicates 88.4% of residues are within the most favoured regions (A, B and L); with 11.6% of residues within additionally allowed regions (a, b, l, p). No residues are within the generously allowed or disallowed regions. (B) PROCHECK statistical analysis of the actual PDB deposited structure indicates 81.4 and 18.6% of residues within most favoured and additionally allowed regions.

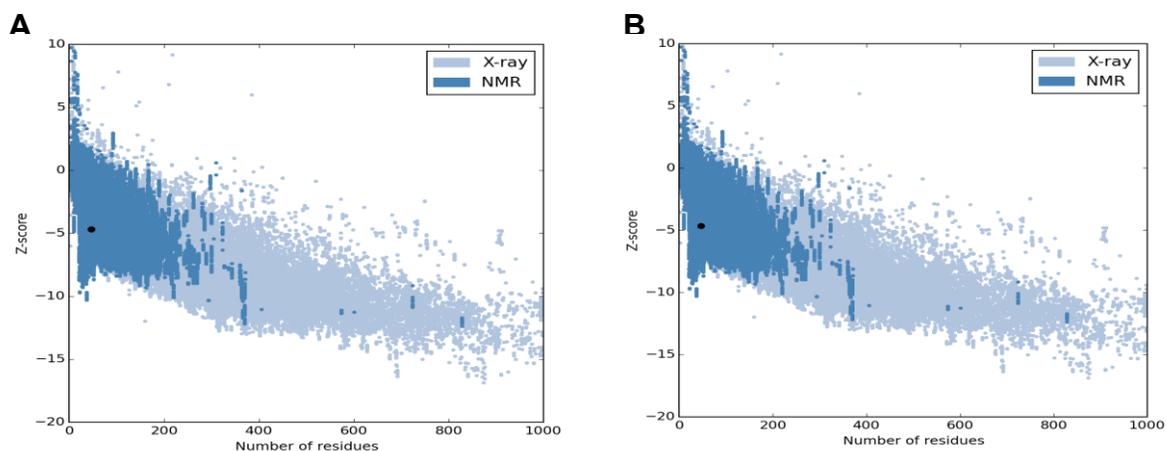


Figure A2.4 Graphical outputs from (A) RNF8 model 641 and (B) deposited structure (PDB: 4AYC) evaluation using ProSa. Z-scores corresponding to NMR derived datasets are depicted by the dark blue colouration, whilst the X-ray crystallographic structural counterparts are shown in light blue. The black dots correspond to Z-score positions of (A) RNF8 model and (B) RNF8 PDB structure, with respect to high-quality datasets.

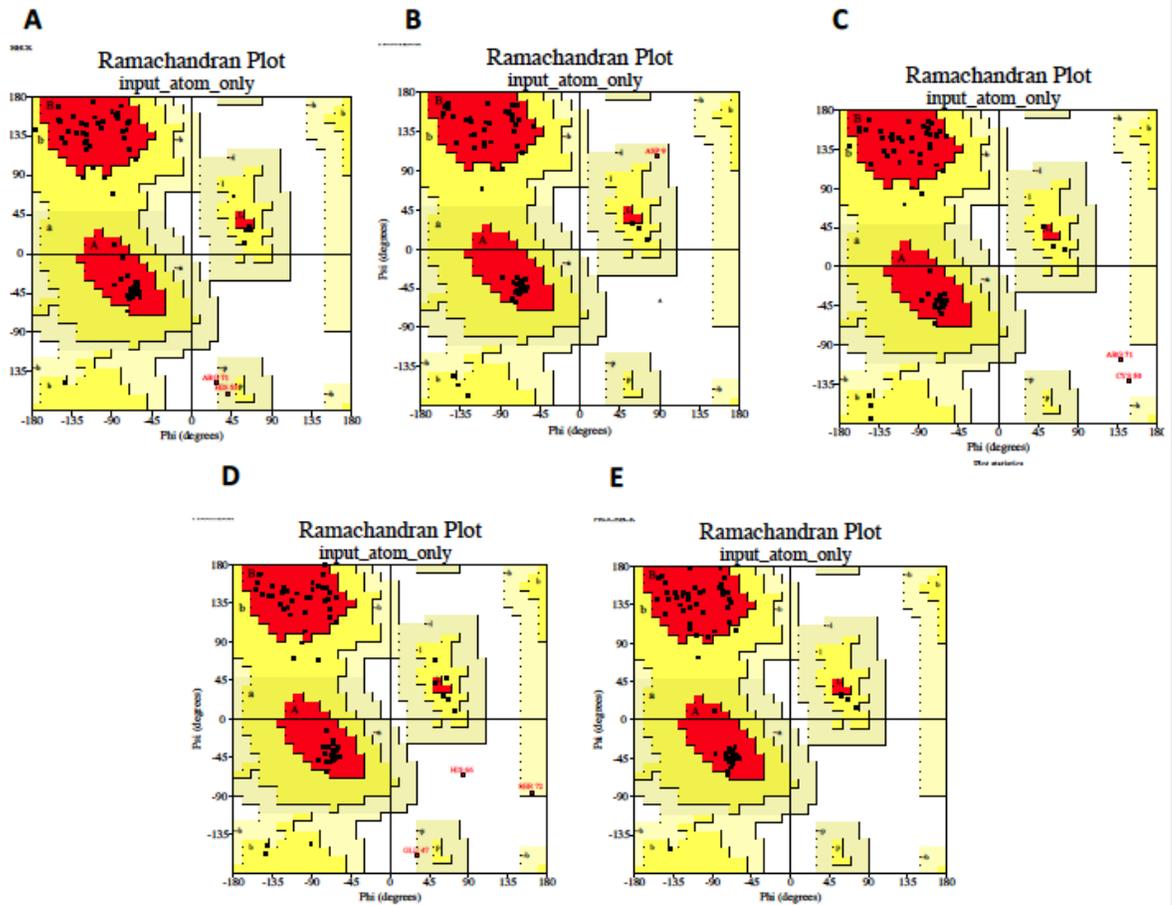


Figure A2.5 Ramachandran plots produced for the CHFR extended RING model via multi-template modelling in MODELLER, corresponding to models (A) 37, (B) 54, (C) 38, (D) 23 and (E) 10.

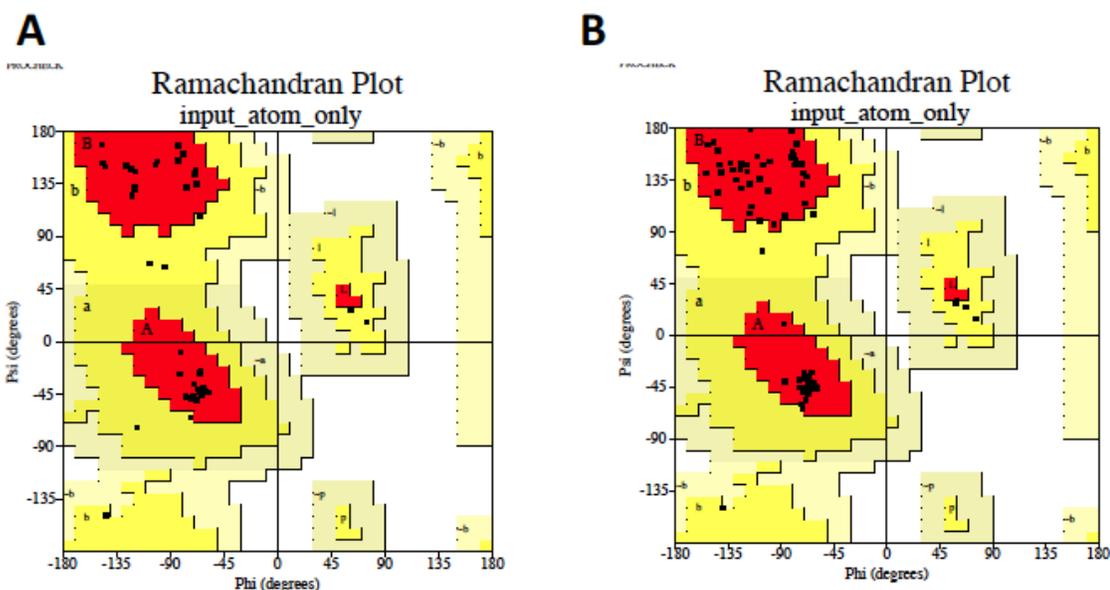


Figure A2.6 Ramachandran plot produced for the CHFR-RING (A) model 106 and (B) (extended) model 10. (A) PROCHECK statistical analysis of model 106 indicates 80.5.0% of residues are within the most favoured regions (A, B and L); with 19.5% of residues within additionally allowed regions (a, b, l, p). No residues are within the generously neither allowed nor disallowed regions for either model. (B) PROCHECK statistical analysis of model 10 indicates 85.0% of residues are within the most favoured regions (A, B and L); with 14.7% of residues within additionally allowed regions (a, b, l, p). No residues are within the generously neither allowed nor disallowed regions for either model.

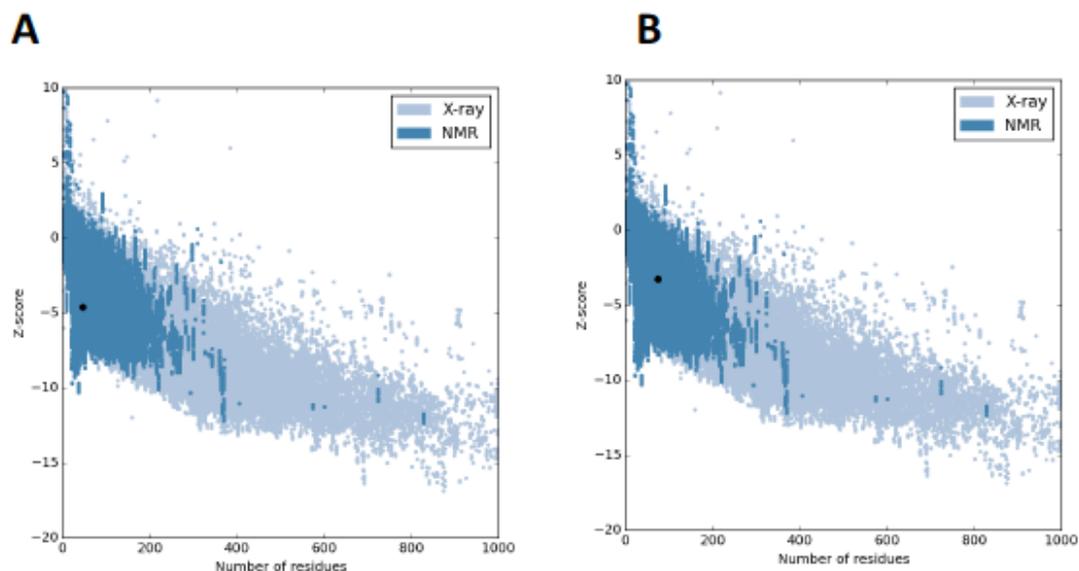


Figure A2.7 Graphical outputs from (A) CHFR RING model 106 and (B) model 10 evaluation using ProSa. Z-scores corresponding to NMR derived datasets are depicted by the dark blue colouration, whilst the X-ray crystallographic structural counterparts are shown in light blue. The black dots correspond to Z-score positions of (A) the RING CHFR model 106 (B) extended model 10, with respect to known high-quality datasets previously described.

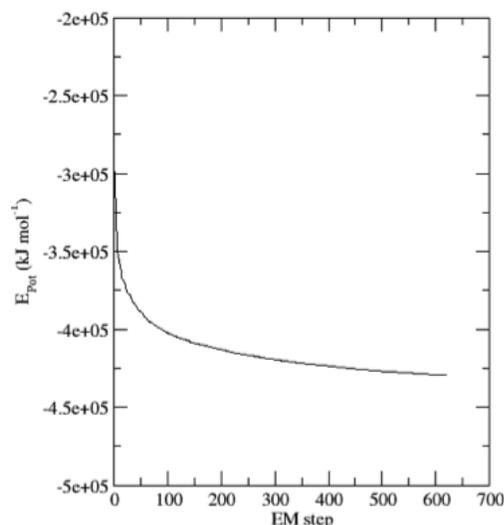


Figure A2.8 Energy minimization (EM) of CHFR RING model 10 (with zinc ions) by the steepest descent algorithm, performed in GROMACS (Van der Spoel *et al.*, 2005). The potential energy (E_{pot}) steadily converges to $<1000\text{kJ mol}^{-1} \text{ nm}^{-1}$ in 677 steps (677ps). Proceeding EM of the system, the average (negative) E_{pot} was $-4.13946 \text{ e}+05 \text{ mol}^{-1} \text{ nm}^{-1}$ and therefore to the order of 10^5 , indicating that the system is stable enough for proceeding simulations.

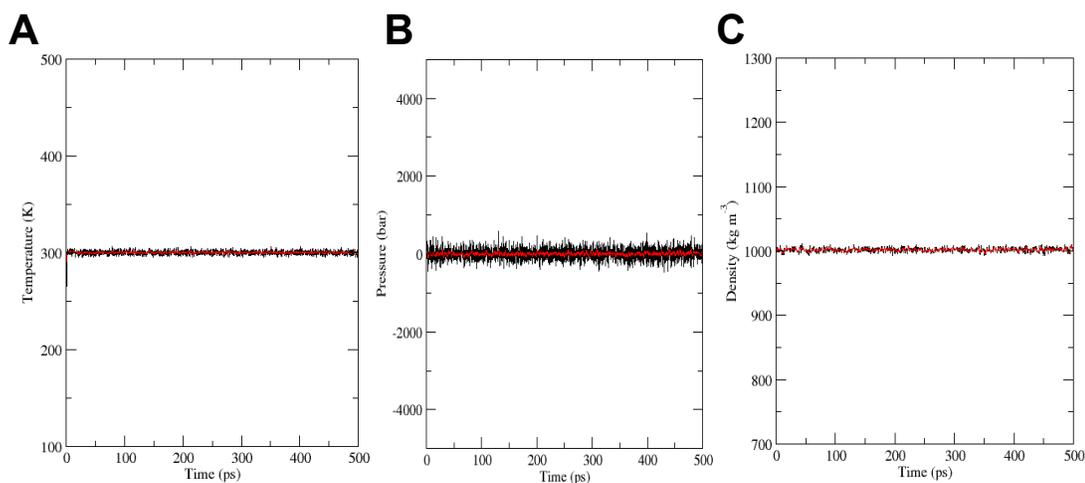


Figure A2.9 CHFR RING model 10 (with zinc ions) trajectories during system equilibration. The CHFR RING equilibration process demonstrates the stability of (A) temperature, (B) pressure and (C) density within the system, for the 500ps simulations. (A) Temperature equilibration included heating the system from an initial temperature of 0 to 300K (within the constant volume) for 500ps and in 250,000 steps. Coordinates, velocities and energies were saved every 0.2 ps for 500 steps. The system temperature quickly reaches the target (300K) and remains close to this value throughout the 500ps equilibration process, with an average temperature of 299.977K. (B) Relaxation of pressure and (C) density coupling were also used, with coordinates, velocities and energies saved every 1.0ps for 500 steps. 10-point running averages are indicated by the red lines (for all three graphs, (A), (B) and (C)), corresponding to overall averages of (B) 3.57597 bar and (C) 1001.83 Kg m^{-3} , respectively.

Appendix 3

Table A3.1 PISA analysis of interactions at the RNF8 (Chain K): Ubc13 (Chain J) interface (Campbell *et al.*, 2012; PDB: 4ORH) compared to the superimposed CHFR RING model 10.

CHFR RING (Ubc13)	RNF8 (Ubc13)	Hydrogen bond?
I305 (R7)	I404 (R7)	Y
I306 (R7)	I405 (R7)	Y
Q308 (R6)	S407 (R6)	Y
D309 (R6)	E408 (R6)	-
W332 (P97)	W430 (P97)	-
P340 (S96)	P438 (S96)	-
R343 (S96)	R441 (S96)	-

Table A3.2 PISA analysis of interactions at the RNF8 (Chain K): Ubc13 (Chain J) interface (Campbell *et al.*, 2012; PDB: 4ORH) compared to the superimposed CHFR RING model 10. Y= yes.

CHFR RING	Ubc13	Hydrogen bond?
Q308	R6	Y
D309		
I305	R7	Y
I306		

Table A3.3 Summary of hydrogen bond distances calculated via PISA analysis of interactions at the RNF8 (Chain K): Ubc13 (Chain J) interface (Campbell *et al.*, 2012; PDB: 4ORH) compared to the superimposed CHFR RING model 10.

CHFR RING (atom) (distance, Å)	RNF8-1 (atom) (distance, Å)	Ubc13 (atom)
R335 (NH2) (1.95)	R433 (NH2) (2.12)	M64 (O)
Q308 (O, OE1) (3.00, 3.00)	S407 (O) (2.55)	R6 (NH2)
C307 (O) (6.10)	C406 (O) (3.71)	R7 (N)
I305 (O) (3.8)	I404 (O) (3.28)	R7 (NH1)
I306 (O) (3.60)	I405 (O) (3.29)	R7 (NH1)
Q308 (3.60)	S407 (OG) 3.33	K10 (NZ)

Table A3.4 PISA analysis of interactions at the RNF4 (Chain A): Polyubiquitin C (Chain E) interface (Branigan *et al.*, 2015; PDB: 5AIU) compared to the superimposed CHFR RING model 10. Y= yes, -= not observed.

RING 10 (Polyubiquitin C)	RNF4 (Polyubiquitin C)	Hydrogen bonds?
E300 (O) (K11, NZ, 3.0)	S196	-/-
T303 (OG1) (K11, NZ, 2.8)	S200	-/-
R343 (NH1) (Q40, OE1, 3.5)	R246 (NH1, O) (Q40, OE1, 2.79) (R72, O, 2.94)	Y/-

Table A3.5 PISA analysis of interactions at the RNF4 (Chain A): Polyubiquitin C (Chain C) interface (Branigan *et al.*, 2015; PDB: 5AIU) compared to the superimposed CHFR RING model 10. Y= yes, -= not observed.

RING 10	Polyubiquitin C (RNF4)	Hydrogen bonds?
H322 (NE2) (E34, O, 2.4)	E34 (O) (H225, NE2, 2.77)	Y/-
R343 (NH1) (Q40, OE1, 3.5)	Q40 (OE1) (R181, NH1, 2.9)	Y/-

Table A3.6 Summary of hydrogen bonds identified via PISA analysis of interactions at the RNF4 (Chain A): Ubc13 (Chain E) interface (Branigan *et al.*, 2015; PDB: 5AIU) compared to the superimposed CHFR RING model 10.

RING 10 (Ubc13)	Ubc13 (RNF4)	Hydrogen bonds? / Salt bridges?
Q308 (O) (K10, NZ, 3.0) Q308 (OE1) (R6, N, 3.1)	R6 (M205*)	Y/Y
I306 (3.2, O) (R7, NH1)	R7 (NH1, NH1) (I138, O, 2.73) (P137, O, 3.71)	Y/-
D309 (OD, OD2) (K10, NZ, 1.5) (K10, NZ, 3.1)	K10 (D206*)	Y/Y
P340 (O) (S96, OG, 2.3)	(S96, OG, 2.6) (P243, O)	Y/-
W332	P97 (S235)	-/-

Table A3.7 Primers used to target and mutate selected residues within the CHFR RING domain. Targeted (and mutated) codons encoding for alanine or cysteine (*) are shown capitalized and in red.

Mutation	Primer pair (5' → 3')
E300A	F: gggaagccagacaagatggag GCG acgctgacatgcatcatctgc R: gcagatgatgcatgtcagcgt CGC ctccatcttgtctggcttccc
T303A	F: gatggaggagacgctg GCA tgcatcatctgccagg R: cctggcagatgatgca TGC cagcgtctcctccatc
I305A	F: gagacgctgacatgc GCC atctgccaggacctg R: caggtcctggcagat GGC gcatgtcagcgtctc
I306A	F: gacgctgacatgcatc GCC tgccaggacctgctgc R: gcagcaggtcctggca GGC gatgcatgtcagcgtc
Q308A	F: gctgacatgcatcatctgc GCG gacctgctgcacgactgcg R: cgcagtcgtgcagcaggtc CGC gcagatgatgcatgtcagc
D309A	F: gcatcatctgccag GCC ctgctgcacgactg R: cagtcgtgcagcag GGC ctggcagatgatgc
H322C*	F: catctgccaggacctgctg TGC gactgctgagtttgcagc R: gctgcaaactcacgcagtc GCA cagcaggtcctggcagatg
W332A	F: gcttgctactcgggc GCG atggagcgcctcgtcc R: ggacgagcgtccat CGC gcccagtagcaagc
R335A	F: cgggctggatggag GCC tcgtccctgtgtcc R: ggacacagggacga GGC ctccatccagcccg
P340A	F: cgctcgtccctgtgt GCT acctgccgctgtccc R: gggacagcggcaggt AGC acacagggacgagcg
R343A	F: ctgtgtcctacctgc GCC tgtcccgtggagcgg R: ccgctccacgggaca GGC gcaggtaggacacag
Y362A	F: ctcaacaacctcgtggaagca GCC ctcatccagcatccagacaag R: cttgtctggatgctggatgag GGC tgttccacgaggtgttgag
L363A	F: caacaacctcgtggaagcatac GCC atccagcatccagacaagagtc R: gactcttgtctggatgctggat GGC gtatgcttccacgaggtgttg

NB: F: forward primer, R: reverse primer.

Table A3.8 Primers used to target and mutate selected residues within ubiquitin.
Targeted (and mutated) codons encoding for alanine are shown capitalized and in red.

Mutation	Primer pair (5' → 3')
E34A	F: ggcaaagatccaagacaag GCA ggcatccctcctgaccagc R: gctggtcaggaggatgcc TGC cttgtcttgatctttgcc
G35A	F: caaagatccaagacaaggaa GCT atccctcctgaccagcagag R: ctctgctggtcaggaggat AGC ttcctgtcttgatctttg
Q40A	F: ggaaggcatccctcctgac GCG cagagggtgatctttgctg R: cagcaaagatcaacctctg CGC gtcaggagggatgccttc
R72A	F: cctgcacctggtcctc GCT ctcagagggtgggtgac R: gtcaccacacctctgag AGC gaggaccaggtgcaggg

NB: F: forward primer, R: reverse primer.

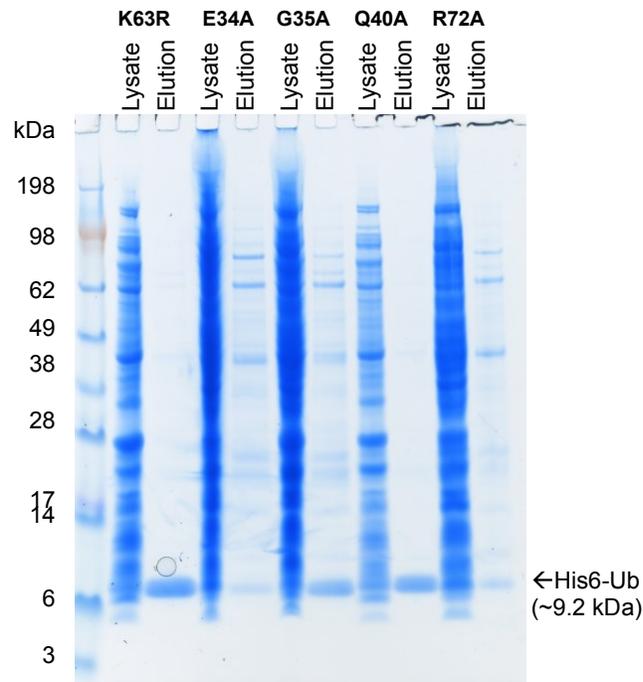


Figure A3.9 Purification of mutant His6-ubiquitin, targeting interactions with the FL-CHFR central RING domain. Mutant ubiquitin proteins (~9.2 kDa) were expressed and purified by metal ion affinity chromatography (HisTrap FF, GE Healthcare), as previously described for all His6-E2 ubiquitin-conjugating enzyme.

CHAPTER 7: Bibliography

Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T., Weng, J. (1987) Protein Data Bank In: Allen, F. H., Bergerhoff, F. H., Sievers, R. (eds), Crystallographic databases – Information, Content, software Systems, *Scientific Applications, Commission of the International Union of Crystallography*, Cambridge, UK.

Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., Lindahl, E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2: 19-25.

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez, J., Billis, K., Girón, C. G., Hourlier, T., Howe, K., Kähäri, A., Kokocinski, F., Martin, F. J., Murphy, D. N., Nag, R., Ruffier, M., Schuster, M., Tang, A., Vogel, J., Searle, S. M. (2016) The Ensembl gene annotation system. *Database*. DOI: 10.1093/database/baw093.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2008) *Molecular Biology of the Cell*. 5th Edition. Garland Science (Taylor & Francis Group), United States of America.

Aleem, E., Arceci, R. J. (2015) Targeting cell cycle regulators in hematologic malignancies. *Frontiers in Cell and Developmental Biology* 3(16): 1-22.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:3389-3402.

Anon (2013) E2Scan Kit Version 2: User Protocol. Version 2.1, July 2013. Ubiquigent, Dundee (Scotland).

Anon (2015) Overview of prediction methods: predict secondary structure (PSIPRED). Seen 10 March 2016. URL: <http://bioinf.cs.ucl.ac.uk/index.php?id=779>

Anon (2016) The Sequence Identity and Similarity (SIAS) server. Available at: <http://imed.med.ucm.es/Tools/sias.html>

Apol, E., Apostovov, R., Berendsen, H. J. C., van Buuren, A., Bjelkmar, P., van Drunen, R., Feenstra, A., Groenhof, G., Kasson, P., Larsson, P., Meulenhoff, P., Murolo, T., Pall, S., Pronk, S., Schulz, R., Shirts, M., Sijbers, A., Tieleman, P., van der Spoel, D. B. H., Lindahl, E. (2010). GROMACS USER MANUAL: Version 4.5.6. Seen 12 December 2015. Available at: <ftp://ftp.gromacs.org/pub/manual/manual-4.5.6.pdf>.

Apweiler, R., Armstrong, R., Bairoch, A., Cornish-Bowden, A., Halling, P. J., Hofmeyr, J. H. S., Kettner, C., Leyh, T. S., Rohwer, J., Schomburg, D. (2010) A large-scale protein-function database. *Nature Chemical Biology* 6(11): 785.

Arnold, K., Bordoli, L., Kopp, J., Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22(2): 195-201.

Bairoch, A., Bougueleret, L. and the Uniprot Consortium (2007) The Universal Protein Resource (UniProt). *Nucleic Acids Research* (35): 193-197.

Bakkenist, C. J., Kastan, M. B. (2003) DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* 421(6922): 499-506.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Hermans, J. (1981) Interaction models of water in relation to protein hydration. *Intermolecular forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry*, B. pp: 331-342, Reidel Publication Company, Holland.

Berg, J. M., Tymoczko, J. L., Stryer, L. (2007) *Biochemistry*. 6th Edition. W. H. Freeman and Company, New York (USA).

Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M., Kim, P. S. (1995) Predicting coiled coils by use SF pairwise residue correlations. *Proceedings of the National Academy of Sciences of the United States of America* 92 (18): 8259-8263.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissi, H., Shindyalov, I. N., Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acid Research* 28: 235-242.

Bothos, J., Summers, M. K., Venere, M., Scolnick, D. M., Halazonetis, T. S. (2003) The Chfr mitotic checkpoint protein functions with Ubc13-Mms2 to form Lys63-linked polyubiquitin chains. *Oncogene* 22(46): 7101-7107.

Branigan, E., Plechanovova, A., Jaffray, E. G., Naismith, J. H., Hay, R. T. (2015) Structural basis for the RING-catalyzed synthesis of K63-linked ubiquitin chains. *Nature Structural and Molecular Biology* 22(8): 597-604.

Brzovic, P. S., Keefe, J. R., Nishikawa, H., Miyamoto, K., Fox, D., Funda, M., Ohtam T., Klevit, R. (2003) Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proceedings of the National Academy of the Sciences of the United States* 100(10): 5646-5651.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. J. (1983) CHARMM- a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* 2: 187-217.

Brzovic, P. S., Lissounov, A., Christensen, D. E., Hoyt, D. W., Klevit, R. W. (2006) A UbcH5/Ubiquitin Noncovalent Complex Is Required for Processive BRCA1-Directed Ubiquitination. *Molecular Cell* 21: 873-880.

Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K., Jones, D. T. (2013) Scalable web service for the PSIPRED protein analysis work bench. *Nucleic Acid Research* 41: 349-357.

Buetow, L., Huang, D. T. (2016) Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nature Reviews Molecular Cell Biology* 17: 626-624.

Bullough, W.S., Johnson, M. (1951) The energy relations of mitotic activity in adult mouse epidermis. *Proceedings of the Royal Society of B Biological Sciences* 138(893): 562-575.

Campbell, S. J., Edwards, R. A., Leung, C. C. Y., Neculai, D., Hodge, C. D., Dhe-Paganon, S., Glover, J. N. M. (2012) Molecular insights into the function of RING finger (RNF)-containing proteins hRNF8 and hRNF168 in Ubc13/Mms2 dependent ubiquitination. *Journal of Biological Chemistry* 287: 23900-23910.

Campisi, J., Medrano, E. E., Morro, G., Pardee, A. B. (1982) Restriction point control of cell growth by a labile protein: evidence for increased stability in transformed cells. *Proceedings of the National Academy of Sciences* 79: 436-440.

Case, D. A., Cheatham III, T. E., Darden, T., Gohlke, H., Luo, R., Merz Jr, K. M., Onufreiv, A., Simmerling, C., Wang, B., Woods, R. J. (2005) The Amber biomolecular simulation programs. *Journal of Computational Chemistry* 26(16): 1668-1688.

Chahwan, R., Gravel, S., Matsusaka, T., Jackson, S. P. (2013) Dma/ RNF8 proteins are evolutionary conserved E3 ubiquitin ligases that target septins. *Cell Cycle* 12(6): 1000-1008.

Chang, L., Zhang, Z., Yang, J., McLaughlin, S. H., Barford, D. (2015) Atomic structure of the APC/C and its mechanism of protein ubiquitination. *Nature* 522: 50-454.

Chaturvedi, P., Sudakin, V., Bobiak, M. L., Fisher, P. W., Mattern, M. R., Jablonski, S. A., Hurle, M. R., Zhu, Y., Yen, T., Zhou, B. S. (2002) Chfr regulates a mitotic stress pathway through its RING-finger domain with ubiquitin ligase activity. *Cancer Research* 62(6): 1797-1801.

- Chazan, A. (Undated) Peptide property calculator. URL:
<http://biotools.nubic.northwestern.edu/proteincalc.html>.
- Chin, C. F., Yeong, F. M. (2010) Safeguarding entry into mitosis: the antephasic checkpoint. *Molecular and Cellular Biology* 30(1): 22-32.
- Cliby, W. A., Roberts, C. J., Cimprich, K. A., Stringer, C. M., Lamb, J. R., Schreiber, S. L., Friend, S. H. (1998) Overexpression of a kinase-inactive ATR protein causes sensitivity to DNA-damaging agents and defects in cell cycle checkpoints. *EMBO Journal* 17(10): 159-169.
- Clamp, M., Cuff, J., Searle, S. M., Barton, G. J. (2002) The Jalview Java alignment editor. *Bioinformatics* 20(3): 426-427.
- Cohen, C., Parry, D. A. D. (1986) α -helical coiled-coils – a widespread motif in proteins. *Trends in Biochemical Sciences* 11(6): 245-248.
- Cohen, C., Parry, D. A. D. (1990) α -helical coiled-coils and bundles: how to design an α -helical protein. *PROTEINS: Structure, Function and Genetics* 7: 1-15.
- Cole, C., Barber, J. D., Barton, G. (2008) The JPred 3 secondary structure prediction server. *Nucleic Acid Research* 36: W197-W201.
- Corn, P. G., Summers, M. K., Fogt, F., Virmani, A. K., Gazdar, A. F., Halazonetis, T. D., El-Deiry, W. S. (2003). Frequent hypermethylation of the 5' CpG island of the mitotic stress checkpoint gene *Chfr* in colorectal and non-small cell lung cancer. *Carcinogenesis* 24(1): 47-51.
- Corabalan, N., Runti, G., Adler, C., Covaceuszach, S., Ford, R. C., Lamba, D., Beius, K., Scozzi, M., Vincent, P. A. (2013) Functional and Structural Study of the Dimeric Inner Membrane Protein SbmA. *Journal of Bacteriology* 195(23): 5352-5361.
- Crick, F., H. C. (1953) The packing of α -helices: simple coiled-coils. *Acta Crystallography* 6: 689-697.
- Cuff, J. A., Barton, G. J. (1999) Evaluation and improvement of multiple sequence methods for secondary structure prediction. *Proteins* 34: 508-519.
- Dalton, J. A. R., Jackson, R. M. (2007) An evaluation of automated homology modelling methods at low target-template sequence similarity. *Structural Bioinformatics* 23(15): 1901-1908.
- Darden, T., York, D., Pedersen, L. (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics* 98: 10089-10092.

Derks, S., Cleven, A. H. G., Melotte, V., Smits, K. M., Brandes, J. C., Azad, N., Crieckinge, V., Bruïne, A. P., Herman, J. G., Engeland, M. (2014) Emerging evidence for CHFR as a cancer biomarker: from tumor biology to precision medicine. *Cancer Metastasis Review* 33: 161-171.

Deshai R.J., Joazeiro, C.A. (2009) RING domain E3 ubiquitin ligases. *Annual Review of Biochemistry* 78: 399-434.

Deveraux, Q., Ustrell, V., Pickart, C. M. (1990) A 25-kilodalton ubiquitin carrier protein (e2) catalyzes multi-ubiquitin chain synthesis via lysine-48 of ubiquitin. *Journal of Biological Chemistry* 265: 21835-21842.

Ding, Z., Lee, G., Liang, X., Gallazzi, F., Arunima, A., van Doren, S. R. (2005) PhosphoTr peptide binding globally rigidifies much of the FHA domain from the *Arabidopsis* receptor kinase-associated protein phosphatase. *Biochemistry* 44: 10119-10134.

Do, M. (1997) Cyclin-dependent kinases: Engines, clocks, and microprocessors. *Annual Review of Cell And Developmental Biology* 13: 261-291.

Drozdetskiy, A., Cole, C., Procter, J., Barton, G. J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acid Research* 43(W1): W389-W394.

Duda, D. M., Borg, L. A., Scott, D. C., hunt, H. W., Hammel, M., Schulman, B. A. (2008) Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. *Cell* 134(6): 995-1006.

Durocher, D., heckle, J., Fersht, A. R., Jackson, S. P. (1999) The FHA domain is a modular phosphopeptide recognition motif. *Molecular Cell* 4: 387-394.

Elledge, S. J. (1996) Cell cycle checkpoints – preventing an identity crisis. *Science* 274: 1664-1672.

Erson, A. E., Petty, E. M. (2004) CHFR-associated early G2/M defects in breast cancer cells. *Molecular Carcinogenesis* 39(1): 26-33.

Estey, M. P., Di Ciano-Oliveira, C., Froese, C. D., Bejide, M. T., Trimble, W.S. (2010). Distinct roles of septins in cytokinesis: SEPT9 mediates midbody abscission. *Journal of Cell Biology* 191:741-749.

Eswarn N., Webb, B., Marti-Renom, M. A., Madhusudhan, M, S., Eramian, D., Shen, M., Sali, A. (2006) Comparative protein structure modelling using modeller, in Baxevanis, A., Davison, D. B., Page, R. D. M., Petsko, G. A., Stormo, G. D. (eds) *Current Protocols in Bioinformatics*, John Wiley and Sons, Inc., New Jersey, 5.6.1-5.6.32.

Fang, G., H. Yu., Kirschner, M. W. (1999) Mitotic Proteolysis. *Philosophical Transactions of the Royal Society B: Biological Sciences* 354:1583–1590.

Evans, P. C., Smith, T. S., Lai, M. J., Williams, M. G., Burke, D. F., Heyninck, K., Kreike, M. M., Beyaert, R., Blundell, T., Kilshaw, P. J. (2003) A novel deubiquitinating enzyme. *Journal of Biological Chemistry* 278(25): 180-186.

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., Chang, H., Dosztányi, S., El-Gebali, S., Fraser, M., Gough, J., Haft, S., Holliday, G. L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marcgeler-Bauer, A., Mi, H., Mistry, J., Natale, D. A., Necci, M., Nuka, G., Orengo, C. A., Park, Y., Pesseat, S., Piovesan, D., Potter, S. C., Rawlings, N. D., Redaschi, N., Piovesan, D., Potter, S. C., Rawlings, N. D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Stilltoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Wu, C. H., Xenarios, I., Yeh, L., Yong, S., Metzchess, A. L. (2017) Interproscan in 2017 – beyond protein family and domain annotations. *Nucleic Acid Research* 45(1): 190-199.

Fischer, H., de Oliveira Neto, M., Napolitano, H. B., Polikarpov, I., Craievich, A. F. (2010) *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* 43: 101–109.

Fiser, A., Do, R. K. G., Sali, A. (2000) Modelling loops in protein structures. *Protein Science* 9: 1753-1773.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., García Girón, C., Gordon, L., Hourlier, T., Hunt, S., Johnson, N., Juettemann, T., Kähäri, A. K., Keenan, S., Kulesha, E., Martin, F. J., Maurel, T., McLaren, W. M., Murphy, D. N., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H. S., Ruffier, M., Sheppard, D., Taylor, K., Thormann, A., Trevanion, S. J., Vullo, A., Wilder, S. P., Wilson, M., Zadissa, A., Aken, B. L., Birney, E., Cunningham, F., Harrow, J., Hubbard, T. J. P., Kinsella, R., Muffato, M., Parker, A., Spudich, G., Yates, A., Zerbino, D. R., Searle, S. M. J. (2014) Ensembl 2014. *Nucleic Acid Research* 42: 749-755.

Franke, D., Svergun, D. I. (2009) *DAMMIF: a program for rapid ab initio shape determination in small-angle scattering*. *Journal of Applied Crystallography* 42: 342-346.

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuykkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M., Svergun, D. I. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography* 50: 1212-1225.

Freemont, P.S., Hanson, I.M., Trowsdale, J (1991) A novel cysteine-rich sequence motif. *Cell* 64:483–84.

Fox, N. K., Brenner, S. E., Chandonia, J. M. (2014) SCOPe: Structural Classification of Proteins – extended, integrating SCOP and ASTRAL data classifications of new structures. *Nucleic Acid Research* 42(1): 304-309.

Fukuda, T., Kondo, Y., Nakagama, H. (2008) The Anti-Proliferative Effects of the CHFR Depend on the Forkhead Associated Domain, but not E3 Ligase Activity Mediated by Ring Finger Domain. *PLoS One* 3(3). DOI: e1776. doi:10.1371/journal.pone.0001776.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook* (Ed: Walker, J. M), Humana Press, USA.

Ginalski, K. (2002) Comparative modeling for protein structure prediction. *Current Opinion in Structural Biology* 16: 172-177.

Glatter, O., Kratky, O. (1982) Small angle X-ray scattering. First Edition. Academic Press, London.

Goodarzi, A. A., Jeggo, P., Lobrich, M. (2010) The influence of heterochromatin on DNA double strand break repair: getting the strong, silent type to relax. *DNA Repair* 9:1273–1282.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, j., Lopez, R. (2010) A new bioinformatics analysis tools framework at EMBL. *Nucleic Acid Research* 38(1): 695-699.

Graewert, M. A., Franke, D., Jeffries, C. M., Blanchet, C. E., Ruskule, D., Kuhle, K., Flieger, A., Schafer, B., Tartsch, B., Meijers, R., Svergun, D. I. (2015). Automated pipeline for purification, biophysical and X-ray analysis of biomacromolecular solutions. *Scientific Reports* 5 (10734). DOI: 10.1038/srep10734.

Guertin, D.A., Venkatram, S., Gould, K. L., McCollum, D. (2002) Dma1 prevents mitotic exit and cytokinesis by inhibiting the separation initiation network (SIN). *Developmental Cell* 3: 779–790.

Guex, N., Peitsch, M. C. (1997) SWISS-MODEL and the SWISS-Pdb viewer: an environment for comparative protein modeling. *Electrophoresis* 18: 2714-2723.

Guinier (1939) La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. *Annales de Physique* 11(12) 161-237.

Jeffries, C. M., Graewert, M. A., Blanchet, C. E., Langley, D. B., Whitten, A. E., Svergun, D. I. (2016) Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nature Protocols* 11(11): 2122-2153.

Hari, K. L., Santerre, A., Sekelsky, J. J., Mckim, K. S., Boyd, J. B., Hawley, R. S. (1995) the mei-41 gene of *Drosophila-melanogaster* is a structural and functional homolog of the human ataxia-telangiectasia gene. *Cell* 82(5): 815-821.

Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* (57): 97-109.

Hershko, A., Ciechanover, A. (1992) The Ubiquitin system for protein degradation. *Annual Reviews in Biochemistry* 61: 761-807.

Hershko, A., Heller, H., Eytan, E., Kaklij, G., Rose, I. A. (1984) Role of the alpha-amino group of protein in ubiquitin-mediated protein breakdown. *Proceedings of the National Academy of the Sciences of the United States* 81(22): 7021-7025.

Hershko, A., Ciechanover, A. (1998) The ubiquitin system. *Annual Reviews in Biochemistry* 67: 425-479.

Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced and scalable molecular simulation. *Journal of Chemical Theory and Computation* 4: 435-447.

Hoffmann K, Bucher P. (1995). The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends in Biochemical Sciences* 20: 347–349.

Hoops, S. C., Anderson, K. W., Merz Jr, K. M. (1991) Force field design for metalloproteins. *Journal of the American Chemical Society* 113: 8262-8270.

Huang, A., Hibbert, R. G., De Jong, R. N., Das, D., Sixma, T. K., Boelens, R. (2011) Symmetry and asymmetry of the Ring-Ring dimer of Rad18. *Journal of Molecular Biology* 410: 424-435.

Hubbard, T. J. P., Blundell, T. L. (1987) Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modeling. *Protein Engineering* 1(3): 156-171.

Huelsenbeck, J. P., Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8): 754-755.

Hurley, J. H., Lee, S., Prag, G. (2006) Ubiquitin-binding domains. *Biochemical Journal* 399(3): 361-372.

Joazerio, C. A. P., Wing, S. S., Huang, H. K., Levenson, J. D., Hunter, T., Liu, Y. C. (1999) The tyrosine kinase negative regulator c-Cbl as a RING –type, E2 dependent ubiquitin-protein ligase. *Science* 286(5438): 309-312.

Johnson, A. E., Gould, K. (2011) Dma1 ubiquitinates the SIN scaffold, Sid4, to impede the mitotic localization of Plo 1 kinase. *EMBO Journal* 30: 341-351.

Johnson, A. E., Coiller, S. E., Oh, M. D., Gould, K. L. (2012) Fission yeast dma1 requires ring domain dimerization for its ubiquitin ligase activity and mitotic checkpoint function. *Journal of Biological Chemistry* 287(31): 25741-25748.

Jones, D. T., Waylor, W. R., Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences.

Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* 292: 195-202.

Jones, P., Binns, D., Chang, H., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sandgrador-Vegas, A., Scheremetjew, M., Yong, S., Lopez, R., Hunter, S. (2014) InterProScan 5: genome scale protein classification. *Bioinformatics* 30(9): 1236-1240.

Jorgensen, W. L.; Chandrasekhar, J., Madura, J. D.; Impey, R. W., Klein, M. L. J. (1983) Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* 79(2): 926-935.

Jorgensen , W. L., Tiradorives, J. (1988) The OPLS potential functions for proteins – energy minimizations for crystals of cyclic-peptides and cambin. *Journal of the American Chemical Society* 110(6): 1657-1666.

Jung, J. W., Bae, S. J., Kang, G., Kim, K., Yeo, W., Park, S., Seol, J. H., Yi, E. C., Kin, K. P. (2013) Analysis of the biochemical role of Lys-11 polyubiquitin chain formation using quantitative mass spectrometry. *Rapid Communications in Mass Spectrometry* 27: 339-346.

Kabir, F. M. L., Alvarez, C. E., Bird, R. C. (2016) Canine mammary carcinomas: a comparative analysis of altered gene expression. *Veterinary Science* 3(1): 1-21.

Kang, D. M., Chen, J., Wong, J., Fang, G. W. (2002) The checkpoint protein Chfr is a ligase that ubiquitinates Plk1 and inhibits Cdc2 at the G2 to M transition. *Journal of Cell Biology* 156(2): 249-259.

Kashima, L., Toyota, M., Mita, H., Suzuki, H., Idogawa, M., Ogi, K., Sasaki, Y., Tokino, T. (2009) CHFR, a potential tumor suppressor, downregulates interleukin-8 through the inhibition of NF-KB. *Oncogene* 28: 2643-2653.

Kashima, L., Iogawa, M., Mita, H., Shitashige, M., Yamada, T., Ogi, K., Suzuki, H., Toyota, M., Ariga, H., Sasaki, Y., Kokino, T. (2012) CHFR protein regulates mitotic checkpoint by targeting PAR-1 protein for ubiquitination and degradation. *Journal of Biological Chemistry* 287: 12975-12984.

Kaufmann, K. W., Lemmon, G. H., DeLuca, S. L., Sheehan, J. H., Meiler, J. (2010) Practically useful: what ROSETTA protein modeling suite can do for you. *Biochemistry* 49: 2987-2998.

Kim, J. M., Cho, E. N., Kwon, Y. E., Bae, S. J., Kim, M., Seol, J. H. (2010) CHFR functions as a ubiquitin ligase for HLTF to regulate its stability and functions. *Biochemical and Biophysical Research Communications* 395: 515-520.

King, R.W., Jackson, P. K., Kirschner, M. W. (1994) Mitosis in transition. *Cell* 79:563–571.

Komander, D., Clague, M. J., Urbe, S. (2009) Breaking the chains: structure and function of the deubiquitinases.

Komander, D. (2010). Mechanism, specificity and structure of the deubiquitinases. *Nature Reviews Molecular and Cell Biology* 10: 550-563.

Konatev, P. V., Petoukhov, M. V., Volkov, V. V., Svergun, D. I. (2006) ATSAS 2.1: a package for small-angle scattering data analysis. *Journal of Applied Crystallography* 39(2): 277-286.

Kopp, J., Schwede, T. (2004) The SWISS-MODEL repository of annotated three-dimensional protein structure homology models. *Nucleic Acid Research* 32: 230-234.

Kratky, O (1982) *Natural high polymers*, in Glatter, O., Kratky, O. Small angle X-ray scattering. Academic Press, London.

Kreyszig (1979) *Advanced Engineering Mathematics*. 4th Edition. Wiley, USA.

Krissinel, A., Henrick, K. (2007) Interface of macromolecular assemblies from a crystalline state. *Journal of Molecular Biology* 372: 774-797.

Kwon, Y. E., Kin, Y. S., Oh, Y. M., Seol, J. H. (2012) Nuclear localization of Chfr is essential for its checkpoint function. *Molecules and cells* 27: 259-363.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993) PROCHECK – a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26(2): 283-291.

Lauffenburger, D. A., Horwitz, A. F. (1996) Cell migration: a physically integrated molecular process. *Cell* 84: 359–369.

Lee, P. S. W., Wang, Y., Dominguez, M. G., Yeung, Y., Murphy, M. A., Bowtell, D. D. L., Stanley, E. R. (1999) The Cbl protooncoprotein stimulates CSF-1 receptor

Lesk, A. M. (2014) Introduction to bioinformatics. 4th Edition. Oxford University Press (UK).

Levkowitz, G., Waterman, H., Ettenberg, S. A., Katz, M., Tsygankov, A. Y., Alroy, I., Lavi, S., Iwai, K., Reiss, Y., Ciechanover, Lipkowitz, S., Yarden, Y. (1999) Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth factor signaling by c-Cbl-Sli-1. *Molecular Cell* 4(6): 1029-1040.

Linke, K., Cace, P. D., Smith, C. A., Vaux, D. L., Silke, J., Day, C. L. (2008) Structure of the Mdm2/ Mdmx RING domain heterodimer reveals dimerization is required for their ubiquitylation *in trans*. *Cell Death and Differentiation* 15(5): 841-848.

Li, P., Merz, K. M. (2015) ZAFF modeling tutorial. Seen 02 January 2016. URL: <http://ambermd.org/tutorials/advanced/tutorial20/ZAFF.htm>

Li, P., Roberts, B. P., Chakravorty, D. K., Merz Jr, K. M. (2013) Rational design of particle mesh ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent. *Journal of Chemical Theory and Computation* 9: 2733-2748.

Liu, C., Wu, J., Paudyal, S. C., You, Z., Yu, X. (2013) CHFR is important for the first wave of ubiquitination at DNA damage sites. *Nucleic Acid Research* 41: 1698-1670.

Lorick, K. L., Jensen, J. P., Weissman, A. M. (2005) Expression, Purification, and Properties of the Ubc4/5 Family of E2 Enzymes. *Methods in Enzymology* 298: 54-68.

Lorick, K. L., Jesen, J. P., Fang, S., Ong, A. M., Hatakeyama, S., Weissman, A. M. (1999) RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proceedings of the National Academy of Sciences* 96: 11364-11369.

Lupas, A., van Dyke, M., Stock, J. (1991) Predicting coiled coils from protein sequences. *Science* 252: 1162-1164.

Lupas, A. N. (Undated) COILS (Version 2.2) Documentation. URL: http://www.ch.embnet.org/software/coils/COILS_doc.html. Seen 28 August 2017.

Lüthy, R., Bowie, J. U., Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356: 83-85.

Madhusudhan, M. S., Marti-Renom, M. A., Sanchez, R., Sali, A. J. (2006) Variable gap penalty for protein structure-sequence alignment. *Protein Engineering, Design and Selection* 19(3): 129-133.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., Simmerling, C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation* 11(8): 3696-3713.

Mahoney, M. W., Jorgensen, W. L. (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, non-polarizable potential functions. *Journal of Chemical Physics* 112(20): 8910-8922.

Marin, I., Ferrus, A. (2002) Comparative Genomics of the RBR family, including the Parkinson's disease-related gene Parkin and the genes of the Ariadne subfamily. *Molecular and Evolution Biology* 19(12): 2039-2050.

Mark, P., Nilsson, L. (2001) Structure and dynamics of the TIP3P, SPC and SPC/E water models at 298 K. *Journal of Physical Chemistry A*: 105(3): 9954-9960.

Marti-Renom, M. A., Madhusudhan, M. S., Sali, A. (2004) Alignment of protein sequences by their profiles. *Protein Science* 13: 1071-1087.

Mattiroli, F., Sixma, T. K. (2014) Lysine-targeting specificity in ubiquitin and ubiquitin-like modification pathways. *Nature Structural and Molecular Biology* 21: 308–316.

Matsusaka, T., J. Pines. (2004) Chfr acts with the p38 stress kinases to block entry to mitosis in mammalian cells. *Journal of Cell Biology* 166: 507–516.

McDonnell, T., Jiang, T., Keating, A. E., Berger, B. (2006) Paircoil2: improved prediction of coiled-coils from sequence. *Bioinformatics* 22(3): 356-358.

Metal Protein Databank (MetalPDB) (2016) MetalPDB statistics- sumamry. [Online] [Accessed 01 April 2016]. Available from:
<http://metalweb.cerm.unifi.it/statistics/summary/>

Metzger, M. M., Prundea, J. N., Klevit, R. E., Weissman, A. M. (2014) RING-type E3 ubiquitin ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitin. *Biochimica et Biophysica Acta* 1843: 47-60.

Merz Jr, K. M (1991) CO₂ binding to human carbonic anhydrase II. *Journal of the American Chemical Society* 113: 406-411.

Middleton, A. J., Day, C. L. (2015) The molecular basis of lysine 48 ubiquitin chain synthesis by Ube2K. *Scientific Reports* 5(16793). DOI: 10.1038/srep16793.

Mikhailov, A., M. Shinohara, C. L. Rieder. (2004) Topoisomerase II and histone deacetylase inhibitors delay the G2 /M transition by triggering the p38 MAPK checkpoint pathway. *Journal of Cell Biology* 166: 517–526.

Mikhailov, A., M. Shinohara., C. L. Rieder. (2005) The p38-mediated stress activated checkpoint. A rapid response system for delaying progression through anaphase and entry into mitosis. *Cell Cycle* 4: 57–62.

Mizuno, K., Osada, H., Konishi, H., Tatematsu Y., Yatabe, Y., Mitsudomi, T., Fujuu, Y., Takahashi, T. (2002) Aberrant hypermethylation of the CHFR prophase checkpoint gene in human lung cancers. *Oncogene* 21(15): 2328-2333.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., Thornton, J. M. (1992) Stereochemical quality of protein structure coordinates. *Proteins: Structure, Function and Genetics* 12: 345-354.

Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., Tramontano, A. (2013) Critical assessment of methods of protein structure prediction (CASP) – round. X. *Proteins: Structure, Function and Bioinformatics* 82(S2):1-6.

Moult, J. (2005) A decade of CASP: progress bottlenecks and prognosis in protein structure prediction. *Current Opinions in Structural Biology* 15: 285-289.

Mueller, P.R., T.R. Coleman, A. Kumagai, Dunphy, W. G. (1995) Myt1: a membrane-associated inhibitory kinase that phosphorylates Cdc2 on both threonine-14 and tyrosine-15. *Science* 270: 86–90.

Murone, M. Simanis, V. T. (1996) The fission yeast *dma1* gene is a component of the spindle assembly checkpoint, required to prevent septum formation and premature exit from mitosis if spindle function is compromised. *EMBO Journal* 15: 6605–6616.

Murray, A.W., Kirschner, M. W. (1989) Cyclin synthesis drives the early embryonic cell cycle. *Nature*: 339:275–280.

Nijman, S. M., Luna-Vargas, M. P., Velds, A., Brummelkamp, T. R., Dirac, A. M., Sixma, T. R., Bernards, R. (2005) A genomic and functional inventory of deubiquitinating enzymes. *Cell* 123(5): 773-786.

Oberoi, J., Richards, M. W., Crumpler, S., Brown, N., Blagg, J., Bayliss, R. (2010) Structural basis of pol(ADP-ribose) recognition by the multizinc binding domain of checkpoint with forkhead-associated and RING domain (CHFR). *Journal of Biological Chemistry* 285(50): 39348-39358.

Oh, Y. M., Kwon, Y. E., Kim, J. M., Bae, S. J., Lee, B. K., Yoo, S. J., Chung, C.H., Deshaies, R. J., Seol, J. H. (2007) Deubiquitination of Chfr, a checkpoint protein, by USP7/ HAUSP regulates its stability and activity. *Biophysical and Biochemical Research Communications* 357: 615-619.

Oh, Y. M., Kwon, Y. E., Kim, J. M., Bae, S. J., Lee, B. K., Yoo, S. J., Chung, C. H., Dashaies, R. J., Seol, J. H. (2009) Chfr is linked to tumour metastasis through the downregulation of HDAC1. *Nature Cell Biology* 11(3): 295-300.

Pardee, A. B. (1974) A restriction point for control of normal animal cell proliferation. *Proceedings of the National Academy of Sciences* 71: 1286-1290.

Pauling, L., Corey, R. B., Branson, H. R. (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America* 37: 205-211.

Pauling, L., Corey, R. B. (1953) Compound helical configurations of polypeptide chains – structure of proteins of the alpha-keratin type. *Nature* 171(4341): 59-61.

Pickart, C. M. (2001) Mechanisms underlying ubiquitination. *Annual Review of Biochemistry* 70: 503-533.

Plechanovova, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H., Hay, R. T. (2012). Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* 489: 115-122.

Porod, G. (1982). *Small angle X-ray Scattering*. Ed: Glatter, O., Kratky, O. Academic Press, London.

Privette, L. M., Weier, J. F., Nguyen, H. N., Yu, X., Petty, E. M. (2008) Loss of CHFR in human mammary epithelial cells causes genomic instability by disrupting the mitotic spindle assembly checkpoint. *Neoplasia* 10(7): 643-652.

Pronk, S., Pail, S., Shultz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., der Spoel, D., Hess, B., Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7): 845-854.

Ptak, C., Varelas, X., Moraes, T., McKenna, S., Ellison, M. J. (2005) Purification and Properties of the Ubiquitin- Conjugating Enzymes Cdc34 and Ubc13-Mms2. *Methods in Enzymology* 298: 43-54.

Putnam, C. D., Hammel, M., Hura, G. L., Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews in Biophysics* 40(3): 191-285.

Páll, S., Abraham, M. J., Kutzner, C., Hess, B., Lindahl, E. (2015) Tackling exascale software challenges in molecular dynamic simulations with GROMACS. Proceedings of the International Conference on Exascale Applications and Software (EASC 2014) 8759: 3-27.

Peters, M. B., Yang, Y., Füsti-Molnár, L., Weaver, M. N., Merz Jr, K. M. M. (2010) Structural survey of zinc-containing proteins and development of the Zinc AMBER Force Field (ZAFF). Journal of Chemical Theory and Computation 6: 2935-2947.

Ramachandran, G. N., Ramakrishnan, C., Sasisekharan, V. (1962) Stereochemistry of polypeptide chain configurations. Journal of Molecular Biology 7: 95-99.

Rambo, R. (2017). SCATTER (Version 1). Available at:
<http://www.bioisis.net/scatter>.

Rambo, R (Undated) Diamond B21: Equipment Description. URL:
<http://www.diamond.ac.uk/Beamlines/Soft-Condensed-Matter/small-angle/B21/description.html>. Seen 14 August 2017.

Retief, J. D. (2000) Phylogenetic Analysis Using PHYLIP. Methods in Molecular Biology 132: 243-258.

Ricke, R. M., van Ree, J. H., van Deursen, J. M. (2008) Whole chromosomal instability in cancer: a complex relationship. Trends in Genetics 24(9): 457-466.

Rohl, C. A., Stauss, C. E. M., Chivian, D., Baker, D. (2004) Modeling structurally variable regions in homologous proteins with Rosetta. Proteins: Structure, Function and Bioinformatics 55: 656-677.

Rose, P. W., Prlic, A., Bi, C., Bluhm, W. F., Christie, C. H., Dutta, S., Green, R. K., Goodsell, D. S., Westbrook, J. D., Woo, J., Young, J., Zardecki, C., Berman, H. M., Bourne, P. E., Burley, S. K. (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acid Research 43: 345-356.

Ronquist, F., Huelsenbeck, J., Teslenko, M. (2011) MrBayes Version 3.2 manual: tutorials and model summaries. Available at:
http://mrbayes.sourceforge.net/mb3.2_manual.pdf. Seen 18 August 2017.

RSCB Protein Data Bank (PDB) (2015) PDB Holdings Breakdown. [Online] [Accessed 11 March 2016]. Available from:
<http://www.rcsb.org/pdb/statistics/holdings.do>

Russell, P., Nurse, P. (1986) Cdc25 functions as an inducer in the mitotic control of fission yeast. Cell 45:145–153.

Sali, A., Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* 234: 779-815.

Sanbhnani, S., Yeong, F. M. (2012) CHFR: a key checkpoint component implicated in a wide range of cancers. *Cellular and Molecular Life Sciences* 69: 1669-1687.

Sanchez, R., Sali, A. (2000) Comparative protein structure modelling. Webster, D. M (Ed.) In: *Protein Structure Prediction: Methods and Protocols*, Humana Press Inc., New Jersey, USA.

Satoh, A., Toyota, M., Itoh, F., Sasaki, Y., Suzuki, H., Ogi, K., Kikuchi, T., Mita, H., Yamashita, T., Kojima, T., Kusano, M., Fujita, M., Hosokawa, M., Endo, T., Tokino, T., Imai, K. (2003) Epigenetic inactivation of CHFR and sensitivity to microtubule inhibitors in gastric cancer. *Cancer Research* 63(24): 8606-8613.

Scheffner, M., Nuber, U., Huibregtse, J.M. (1995) Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* 373:81–83.

Schreiber, V., Dantzer, F., Ame, J. C., de Murcia, G. (2006) Poly(ADP-ribose): novel functions for an old molecule. *Nature Reviews Molecular Cell Biology* 7(7): 517-528.

Schoichet, B. K., Baase, W. A., Kuroki, R., Matthews, B. W. (1995) A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences of the United States of America* 92: 452-456.

Schlick, T. (2010) *Molecular dynamics: basics*. In: Antman, S. S., Marsden, J. E., Sirovich, L (eds.). In: *Molecular Modelling and Simulation: An Interdisciplinary Guide*, Springer Science and Business Media LLC, New York, USA.

Scolnick, D. M., Halazonetis, T. D. (2000) Chfr defines a mitotic stress checkpoint that delays entry into metaphase. *Nature* 406(6794): 430-435.

Sheng, Y., Hong, J. H., Doherty, R., Srikumar, T., Shiouh, J., Avvakumov G. V., Walker, J. R., Xue, S., Neculai, D., Wan, J. W., Kim, S. K., Arrowsmith, C. H., Raught, B., Dhe-Paganon, S. (2012) A human ubiquitin conjugating enzyme (E2)-HECT E3 ligase structure-function screen. *Molecular and Cellular Proteomics* 11: 329-341.

Sheng, Y., Hong, J. H., Doherty, R., Srikumar, T., Shloush, J., Avvakumov, G. C., Walker, J. R., Xue, S., Neculair, D., Wan, J. W., Kim, S. K., Arrowsmith, C. H., Raught, B., Dhe-Paganon, S. (2012) A human ubiquitin conjugating enzyme (E2)-HECT E3 ligase structure-function scree. *Molecular and Cellular Proteomics* 11(8): 329-341.

Shibata, Y., Naruki, N., Kuwabara, Y., Ishiguro, H., Shidona, N., Sato, A., Kimura, M., Koymama, H., Toymama, T., Nishiwakim T., Kudo, J., Terashita, Y., Konishi, S., Sugiura, H., Fujii, Y. (2002) Chfr expression is downregulated by CpG island hypermethylation in esophageal cancer. *Carcinogenesis* 23(10): 1695-1699.

Schrödinger, LLC (2015) MacPymol: Pymol, enhanced for Mac OS X (Version 1.8.0.5) [Software] Available from:
<https://www.pymol.org/>

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7: 539. DOI: i:10.1038/msb.2011.75.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7: 539-454.

Skou, S., Gillilan, R. E., Ando, N. (2014) Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nature Protocols* 9(7): 1727-1739.

da Silva, S. A. W., Vranken, W. F. (2012) ACPYPE – AnteChamber Python Parser interface. *BMC Research Notes* 5: 367.

Sippl, M. J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function and Genetics* 17: 355-362.

Stavridi, E. S., Huyen, Y., Loreto, I. R., Scolnick, D. M., Halazonetis, T. D., Pavellitch, N. P., Jeffrey, P. (2002) The crystal structure of the Chfr mitotic checkpoint protein and its complex with tungstate. *Structure* 10: 891-899.

Stavridi, E. S., Huyen, Y., Loreto, I. R., Scolnick, D. M., Halazonetis, T. H., Pavletich, N. P., Jeffrey, P. D. (2002) Crystal Structure of the FHA Domain of the Chfr Mitotic Checkpoint Protein and Its Complex with Tungstate. *Structure* 10: 891-899.

Stewart, M. D., Ritterhoff, T., Klevit, R. E., Brzovic, P. S. (2016) E2 enzymes: more than just middle men. *Cell Research* 26: 423-440.

Sturner, E., Behl, C. (2017) The Role of the Multifunctional BAG3 Protein in Cellular Protein Quality Control and in Disease. *Frontiers in Molecular Neuroscience* 10 (177). DOI: 10.3389/fnmol.2017.00177.

Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules in solution using simulated annealing. *Biophysical Journal* 76: 2879-2886.

Svergun D.I., Barberato C. and Koch M.H.J. (1995) CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates *Journal of Applied Crystallography* 28: 768-773.

Svergun, D. I., Koch, M. H. J. (2003). Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics* 66: 1735–1782.

Svergun, D. I., Koch, M. H. J., Timmins, P. A., May, R. P. (2013) *Small angle X-ray and neutron scattering from solutions of biological macromolecules*. First Edition. Oxford University Press, England.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *Annals of Statistics* 22: 1701-1762.

Toyota, M., Sasaki, Y., Satoh, A., Ogi, K., Kikuchi, T., Suzuki, H., Mita, H., Tanaka, N., Itoh, F., Issa, J. P. J. (2003) Epigenetic inactivation of CHFR in human tumours. *Proceedings of The National Academy of Sciences of The United States of America* 100(13): 7818-7823.

Tramontano, A. (1998) Homology modelling with low sequence identity. *METHODS: A Companion to Methods in Enzymology* 14: 293-300.

Tsai, M. D. (2002) FHA: a signal transduction domain with diverse specificity and function. *Structure* 10(7): 887-888.

Tsui, V., Case, D. A. (2001) Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* 56(4): 275-291.

Turner, P. (2000) xmgrace Version 5.1.23. Available from:
<http://plasma-gate.weizmann.ac.il/Grace/>

UniProt/TrEMBL (2016) UniProtKB/TrEMBL Protein Database Release 2016_02 Statistics. [Online] [Accessed 11 March 2016]. Available from:
<http://www.ebi.ac.uk/uniprot/TrEMBLstats>

Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., Berendsen, J. C. (2005) GROMACS: fast, flexible and free. *Journal of Computational Chemistry* 26: 1701-1718.

Vega, C., Abascal, J. L. F., Conde, M. M., Aragines, J. L. (2009) What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday Discussions* 141: 251-276.

Venclovas, C., Margelevicius, M. (2005) Comparative modelling in CASP6 using consensus approach to template selection, sequence-structure alignment and structure assessment. *PROTEINS: Structure, Function and Bioinformatics* 7: 99-105.

Venclovas, C., Zemla, A., Fidelis, K., Moult, J. (1997) Criteria for evaluating protein structures derived from comparative modelling. *Proteins: Structure, Function and Genetics* 1: 7-13.

- Versele, M., Thorner, J. S. (2005) Some assembly required: yeast septins provide the instruction manual. *Trends in Cell Biology* 15: 414-424.
- Volkov, V. V., Svergun, D. I. (2003) Uniqueness of *ab-initio* shape determination in small-angle scattering. *Journal of Applied Crystallography* 36: 860-864.
- Wang, H. B., Wang, L. J., Erdjument-Bromage, H, Vidal, M., Tempst, P., Jones, R. S., Zhang, Y. (2004) Role of histone H2A ubiquitination in polycomb silencing. *Nature* 431(7010): 873-878.
- Wang, B., and Elledge, S.J. (2007) Ubc13/Rnf8 ubiquitin ligases control foci formation of the Rap80/Abraxas/Brca1/Brcc36 complex in response to DNA damage. *Proceedings of the National Academy of the Sciences of the United States* 104: 20759–20763.
- Wang, C., Deng, L., Hong, M., Akkaraju, G. R., Inoue, J., Chen, Z. J. J. (2001) TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* 412(6844): 346-351.
- Wang, W. Y., Malcolm, B. A. (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuickChange™ site-directed mutagenesis. *Biotechniques* 26(4): 680-682.
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., Barton, G. J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9): 1189-1191.
- Wilkinson, K. D., Tashayev, V. L., O'Connor, L. B., Larsen, C. N., Kasperek, E., Pickart, C. M. (1995) Metabolism of the polyubiquitin degradation signal - structure, mechanism, and role of Isopeptidase-T. *Biochemistry* 34(44): 14353-14546.
- Wilks, S. S. (1938) The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Annals of Mathematical Statistics*, 9: 60–62.
- Wright, J. D., Mace, P. D., Day, C. L. (2016) Secondary ubiquitin-RING docking enhances Arkaidia and Ark2C E3 ligase activity. *Nature Structural and Molecular Biology* 23(1): 45-54.
- Wu, J., Chen, Y., Lu, L., Wu, Y., Paulson, M. T., Ljungman, M., Ferguson, D. O., Yu, X. (2011) Chfr and RNF8 synergistically regulate ATM activation. *Nature Structural and Molecular Biology* 18(7): 761-768.
- Ye, Y., Rape, M. (2009) Building ubiquitin chains: E2 enzymes at work. *Nature Molecular and Cell Biology* 10: 755-765.

Yu, X., Minter-Dykhouse, K., Malureanu, L., Zhao, W. M., Zhang, D., Merkle, C. J., Ward, I. M., Saya, H., Fang, G., van Deursen, J., Chen, J. (2005) Chfr is required for tumor suppression and Aurora A regulation. *Nature Genetics* 37: 401–406.

Yoshikawa, S., Arai, R., Murayama, K., Imai, Y., Takahashi, R., Shirouzu, M., Yokoyama, S. (2006) Structure of human ubiquitin-conjugating enzyme E2 G2 (UBE2G2/UBC7). *Acta Crystallography Section F* 62: 330-334.

Yuan, L., Lv, Z., Atkinson, J. H., Olsen, S. K. (2017) Structural insights into the mechanism and E2 specificity of RBR E3 ubiquitin ligase HHARI. *Nature Communications* 8(211). DOI: 10.1038/s41467-017-00272-6.

Yu, X., Minter-Dykehouse, K., Malureanu, L., Zhao, W., Zhang, D., Merkle, C. J., Ward, I. M., Saya, H., Fang, G., van Dersen, J., Chen, J. (2005) Chfr is required for tumour suppression and Aurora A regulation. *Nature Genetics* 37(4): 401-406.

Zdobnov, E. M., Apweiler, R. (2001) InterProScan – an integrated platform for the signature recognition methods in InterPro. *Bioinformatics* 17(9): 847-848.

Zdobnov, E. M., Apweiler, R. (2001) InterProScan 5. Available at:
<http://www.ebi.ac.uk/interpro/>. Seen: 20 September 2017.

Zhang, Z., Kulkarni, K. A., Hanrahan, S. J., Thompson, A. K., Barford, D. (2010) The Apc/C Subunit Cdc16/Cut9 is a Contiguous Tetratricopeptide Superhelix with a Homo-Dimer Interface Similar to Cdc27. *EMBO Journal* 29(21): 3733-3744.

Zvelebil, M. J., Barton, G. J., Taylor, W. R., Stenberg, M. J. E. (1987) Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *Journal of Molecular Biology* 195(4): 957-961.