The structure, dynamics and interactions of the von Willebrand factor C3 domain

Submitted to University College London for the degree of

Doctor of Philosophy

Harrison Edward Rowe O'Brien

May 2021

Institute of Structural and Molecular Biology

Declaration

I, Harrison O'Brien, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Von Willebrand factor (VWF) is a multimeric haemostatic protein comprised of a series of repeat domains. It is responsible for platelet adhesion to the site of vessel injury, binding to collagen exposed by injury and to platelets by glycoprotein Ibα (GPIbα). It also acts as a carrier for coagulation Factor VIII (FVIII), increasing its half-life in circulation and localising activated FVIII to the site of injury. Synthesised primarily in the endothelial cells, VWF multimers are either released constitutively into the bloodstream or ultra-large multimers are stored in the endothelial cell cytoplasm in cylindrical storage granules, Weibel-Palade bodies. The importance of VWF function on maintaining normal haemostasis is demonstrated by the quantitative and qualitative defects exhibited by von Willebrand disease (VWD), the most common inherited bleeding disorder in the world. Type 1 VWD is most common, caused by specific mutations throughout the multimeric protein and characterised by reduced levels of circulating VWF. Of the 6 C-terminal C domains of VWF, C3 contains the most VWD Type 1-causing mutations. Whilst the molecular structures of other domains which comprise the collagen, GpIbα and FVIII binding sites are well characterised, the C domain structures remain largely unsolved. This thesis presents the structure of the VWF C3 domain solved using nuclear magnetic resonance spectroscopy and the effect of specific disulphide bond mutations on the domain structure. The C3 domain is comprised of two distinct subdomains – SD1 and SD2. SD1 is dominated by two β-sheet structures whilst SD2 comprises a single β-sheet structure. Domain stability is maintained by five disulphide bonds, two in each subdomain and the inter-subdomain relationship preserved by a single disulphide bond connecting the two. Mutations of each individual disulphide bond resulted in incomplete folding of the domain, highlighting their importance in maintaining the structural integrity of the C3 domain.

Impact Statement

The findings of this thesis can provide the basis to develop a much clearer picture of the role of the C domains in von Willebrand factor (VWF) synthesis, storage and functionality. The C3 domain structure presented in this case refers to the structure found in multimeric VWF in circulation. This does not account, however, for the interaction of C3 with its neighbouring domains in circulation or under shear stress and the results of this thesis would allow the progression onto such investigations. Whilst the role of the C domains in the VWF bouquet structure during the storage of VWF multimers is well established, the underlying dimeric molecular structure between alike domains remains poorly understood. Through solving the structure of the C3 domain when in circulation and by inducing the onset of C3 dimers using Weibel-Palade body conditions, an understanding can develop of the structural changes that occur between VWF storage and secretion into the bloodstream. Furthermore, with the developing significance of VWF free thiols and the fact that the majority of free thiol sites are located in the C3 domain, this work provides the basis to understanding the effect and subsequent importance of free thiols on VWF functionality. The VWF C4 domain has been solved, however, due to its significant differences to the VWF Type C (VWFC) domain family, it is classed as a VWFC-like domain. With the C3 domain being the first VWFC domain structure within VWF to be solved, this provides the basis to solve the structures of alike domains.

A number of mutations in the C3 domain have also been shown to cause Type 1 VWD, characterised by reduced levels of circulating VWF. With each of these mutations being either of cysteine residues or of alternative residues to a cysteine, this likely causes domain misfolding, leading to issues with intracellular transport and retention within the cell. The work in this thesis highlights the extensive impact of removing a single disulphide bond on domain folding, something which has never been shown on a molecular level. It can form the basis to understand the subsequent issues with intracellular transport and storage of VWF multimers. Overall, the work presented in this thesis can progress onto a greater understanding of domain misfolding and, latterly, pave the way for new interventions and therapies for the treatment of Type 1 VWD.

Acknowledgments

Firstly, I would like to thank my primary supervisor, Prof D. Flemming Hansen, for his endless support throughout this PhD. This was not just confined to the PhD but his understanding of life outside of the lab, for that I am eternally grateful.

I would also like to thank my secondary supervisor and Thesis Chair, Prof Andres Ramos and Dr Andrew Osbourne, for their insight and guidance along the way.

A particular thank you goes to Dr Angelo Figueiredo for his constant help and teaching in NMR. A man always doing whatever he can to help everyone, his work does not go unnoticed and this PhD would have been ten times more difficult without his help. Not to mention someone you can always rely on for a trip to the pub!

To my lab mates over the course of the last three and a half years, you have all made it such an enjoyable time; Dr Vaibhav Shukla and Dr Gogulan Karunanithy are fantastic post docs and never failed to give me a steer in the right direction; Dr Lucas Siemons and Dr Ruth Pritchard who were great fellow PhD students but the office has been far too quiet since they moved on; and Harry Mackenzie, Dr Johannes Kremser and Martin Rosellen, some of the most intelligent yet funniest people I have met – our epic squash games will never be matched. I feel very lucky to have made friends for life and I still hope to be invited back for future Christmas parties despite my habit of making them significantly more expensive!

I would like to thank my family, particularly my three brothers, Tom, Jordan and Elliott, who are my closest friends and I thank you all for supporting me over the last three years. From listening to me ramble about my project to picking up the bill whenever we meet up, I promise that dinner will be on me very soon.

To my gorgeous girlfriend, Maya, we have spent almost three years living on opposite sides of the world yet Maya has stood by me all the way through. We have been counting down to the moment we could finally enjoy our lives together and I could not be happier knowing that moment is coming so very soon.

Finally, I would like to thank my wonderful parents, John and Shirley. I cannot begin to describe how much you have helped and supported me and I know that I could not have got this far without you.

Abbreviations

ADAMTS13	A disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13
Ang-2	Angiopoietin-2
AP-*x*	Activator protein-*x*
BMP	Bone morphogenetic protein
cAMP	Cyclic AMP
CCN3	Cell communication network factor 3
CCPN	Collaborative computing project for NMR
Cdc42	Cell division cycle 42
CK	Cystine knot
Col2A	Collagen IIa
COSY	Correlation spectroscopy
Cryo-EM	Cryogenic electron microscopy
CV2	Crossveinless-2
CYANA	Combined assignment and dynamics algorithm for NMR applications
DDAVP	Desmopressin
DNA	Deoxyribonucleic acid
DOSY	Diffusion ordered spectroscopy
Dsb	Disulphide bond isomerase
DSS	Sodium trimethylsilylpropanesulfonate
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EK	Enterokinase
EM	Electron microscopy
ER	Endoplasmic reticulum
FPLC	Fast protein liquid chromatography
FuDA	Function and data analysis
F*x*	Factor *x*
F*x*a	Activated Factor *x*
GDP	Guanosine diphosphate
Gor	Glutaredoxin/glutathione
Gp	Glycoprotein
Grx1	Glutaredoxin-1
GSH	Reduced glutathione
GSSG	Oxidised glutathione
GTP	Guanosine triphosphate
His_6	6 x Histidine
HSQC	Heteronuclear single quantum coherence
ICAM	Intracellular adhesion molecule
IPAP	In-phase/anti-phase
LB	Luria-Bertani
LRP1	Lipoprotein receptor-related protein-1
MD	Molecular dynamics
MIDAS	Metal ion-dependent adhesion site
mRNA	Messenger RNA
MWCO	Molecular weight cut off

MyRIP	Myosin VIIa and RAB interacting protein
NEB	New England BioLabs
Ni-NTA	Nickel-nitrotriacetic acid
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
OD_{600}	Optical density at a wavelength of 600 nm
PDI	Protein disulphide isomerase
PL	Phospholipid
PSGL	P-selectin glycoprotein ligand
PSVS	Protein structure validation software
Rab	Ras-associated binding
RANK	Receptor activator of nuclear factor kappa B
RANKL	Receptor activator of nuclear factor kappa B ligand
Rap1	Ras-related protein 1
RCI	Random coil index
RDC	Residual dipolar coupling
RGD	Arginine-glycine-aspartic acid
RMSD	Root mean square deviation
SD	Subdomain
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
Slp4-a	Synaptotagmin-like protein 4-a
SP	Signal peptide
TAE	Tris-acetate-EDTA
TCEP	Tris (2-carboxyethyl) phosphine
TEV	Tobacco etch virus
TF	Tissue factor
TFPI	Tissue factor pathway inhibitor
Tie	Tyrosine kinase with immunoglobulin-like loops and epidermal growth factor homology
TIL	Trypsin-inhibitor-like
T_m	Melting temperature
TOCSY	Total correlation spectroscopy
tPA	Tissue plasminogen activator
Trx	Thioredoxin
TSP	Thrombospondin
UL VWF	Ultra-large VWF
UV	Ultraviolet
VCAM	Vascular adhesion molecule
VE	Vascular endothelium
VEGFR	Vascular endothelial GF receptor
VWD	von Willebrand disease
VWF	von Willebrand factor
VWFC	von Willebrand factor type C
VWFD	von Willebrand factor type D
WPB	Weibel Palade body
WT	Wild-type

Units of measurement

Å	Angstrom
b	Base
AU	Absorbance units
°C	Celsius degrees
Da	Daltons
g	Grams
Hz	Hertz
K	Kelvin
L	Litre
M	Molar
ppm	Parts per million

Measurement prefixes

M	Mega
k	Kilo
m	Milli
µ	Micro
n	Nano

Element isotopes

^{12}C	Carbon – mass number 12
^{13}C	Carbon – mass number 13
^{14}N	Nitrogen – mass number 14
^{15}N	Nitrogen – mass number 15

Amino acid abbreviations

Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamic acid	E	Glu
Glutamine	Q	Gln
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Lysine	K	Lys
Leucine	L	Leu
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

Contents

Declaration 2
Abstract 3
Impact Statement 4
Acknowledgments 5
Abbreviations 6
Contents 9
List of Figures 12
List of Tables 15
1. Introduction 16
1.1. Haemostasis and the Coagulation Cascade 16
1.2. Von Willebrand Factor 19
1.3. VWF Genetics and translation product 20
1.4. Primary VWF functions and binding sites 22
1.4.1. D domains 23
1.4.2. A domains 26
1.4.3. C domains 30
1.5. Life Cycle of VWF 35
1.5.1. Dimerisation in the endoplasmic reticulum 35
1.5.2. Glycan Modifications in the Golgi network 37
1.5.3. Multimerisation in the Golgi 38
1.5.4. Storage and secretion 40
1.5.5. Platelet-derived VWF 47
1.5.6. Multimer distribution and VWF self-association 48
1.5.7. VWF clearance 51
1.6. Secondary functions of VWF 54
1.6.1. Inflammation 54
1.6.2. Angiogenesis 56
1.6.3. Tumour metastasis 57
1.7. Von Willebrand Disease 59
1.7.1. Type 2 VWD 59
1.7.2. Type 3 VWD 61
1.7.3. Type 1 VWD 61

1.8. Structure determination using NMR spectroscopy 63
1.8.1. Chemical shift 63
1.8.2. J-Couplings 65
1.8.3. NMR experiments and chemical shift assignment 65
1.8.4. Nuclear Overhauser effect 69
1.9. Project aim 70
2. Results: Expression, purification and NMR backbone assignment of VWF C3 71
2.1. VWF C3 (FXa) construct form, expression and purification 73
2.2. VWF C3 purification and NMR analysis 77
2.3. Isotopically-labelled VWF C3 expression and purification 82
2.4. NMR triple resonance experiments and backbone amide assignment 86
2.5. Partial backbone assignment and linker impurity 91
2.6. VWF C3 (EK) plasmid design and expression 93
2.7. Cleavage optimisation, purification and NMR analysis 95
2.8. Isotopically-labelled expression and NMR assignment completion 99
3. Results: Three-dimensional structure of the VWF C3 domain 102
3.1. Assignment of the aliphatic ^{1}H and ^{13}C side chain chemical shifts 102
3.2. Secondary structure assignment 110
3.3. Disulphide bonds and tertiary structure assignment 113
3.4. Residual Dipolar Couplings 126
3.5. High Resolution Structure of VWF C3 127
3.6. Structural comparison to VWFC domains 132
3.7. Structural comparison to the VWF C4 domain 135
3.8. Interaction between the VWF C3 and C4 domains 137
3.9. Expression of the C3L plasmid 141
4. Results: The effect of cysteine mutations on the structure of VWF C3 145
4.1. Mutations of the disulphide bonds of SD1 147
4.2. Mutations of the disulphide bonds of SD2 152
4.3. Mutation of the SD1-SD2 interconnecting disulphide bond 155
5. Concluding remarks 159
6. Materials and Methods 165
6.1. LB expression and purification of TrxA-His_6-VWFC3-His_6 165
6.1.1. Protein Purification 166
6.1.2. FXa cleavage and purification 167
6.2. TrxA-His_6-VWFC3 (FXa) construct 167
6.2.1. Nonsense mutation of the TrxA-His_6-VWFC3-His_6 construct 167

6.2.2. Expression of TrxA-His_6-VWFC3 in SHuffle *E. Coli* cells 170
6.2.3. Purification of the VWF C3 protein 171
6.2.4. Ellman's assay 172
6.2.5. NMR experimental setup 173
6.2.6. Diffusion experiment 174
6.3. TrxA-His_6-VWFC3 (EK) construct 174
6.3.1. New Plasmid design, protein expression and purification 174
6.3.2. NMR experiments and chemical shift assignment 176
6.3.3. Talos+ prediction of backbone dihedral angles and order parameter S^2 177
6.3.4. NOE assignment 177
6.3.5. Residual dipolar couplings and relaxation experiments 178
6.3.6. CYANA 3.0 structure calculation 178
6.4. VWF C4 domain expression, purification and NMR analysis 179
6.5. Expression, purification and NMR analysis of the C3L protein 180
6.6. Disulphide bond mutations of the VWF C3 domain 180
7. Appendix 184
7.1. VWF C3 chemical shift assignment 184
7.2. Ramachandran plot of twenty lowest energy structures of VWF C3 189
8. References 190

List of Figures

Figure 1.1: Overview of the coagulation cascade 17
Figure 1.2: Monomeric VWF domain assignment 20
Figure 1.3: Binding sites of the mature VWF monomer 23
Figure 1.4: Structure of D'D3 region 24
Figure 1.5: Structure of VWF A1 and A3 domains 27
Figure 1.6: Structure of the VWF C4 domain 31
Figure 1.7: Structures of VWFC domains of Col2A, CCN3 GF and CV2 33
Figure 1.8: The major N-linked glycan structure of VWF 36
Figure 1.9: O-linked glycan structures 37
Figure 1.10: Mapping the N- and O-linked glycans in the VWF monomer 38
Figure 1.11: Dimerisation and multimerisation of VWF 40
Figure 1.12: EM studies of VWF A3-CK bouquet structure 42
Figure 1.13: VWF bouquet structure and multimerisation 42
Figure 1.14: WPB formation and VWF secretion pathways 46
Figure 1.15: Mapping of the VWF free thiols 50
Figure 1.16: Structure of the VWF A2 domain 52
Figure 1.17: The role of VWF in inflammation 55
Figure 1.18: Energy levels, free induction decay and Fourier transform 64
Figure 1.19: Biosynthesis of valine and leucine residues using fractional ^{13}C-labelling 68
Figure 2.1: VWF C3 sequence and predicted disulphide bonds 71
Figure 2.2: VWF C3 purification plan 74
Figure 2.3: Post-nonsense mutation construct sequence 75
Figure 2.4: Ni-NTA purification of the unlabelled TrxA-His$_6$-VWFC3 and cleavage optimisation 76
Figure 2.5: ^{15}N-labelled VWF C3 and TrxA-His$_6$ purification 77
Figure 2.6: ^{15}N-HSQC overlay of VWF C3, TrxA-His$_6$ and the uncleaved TrxA-His$_6$-VWFC3 79
Figure 2.7: Stability of VWF C3 over a 2-week time course 81
Figure 2.8: Isotopically-labelled VWF C3 purification 82
Figure 2.9: ^{15}N-HSQC of the isotopically-labelled VWF C3 83
Figure 2.10: ^{15}N-HSQC of reduced VWF C3 85
Figure 2.11: Spin system strips in the backbone assignment spectra 87
Figure 2.12: Sequential assignment using the HNCACB and HN(CO)CACB spectra 88

Figure 2.13: Sequential assignment using the HNCO/HNCACO experiments 90
Figure 2.14: Partial backbone amide assignment of the VWF C3 domain 91
Figure 2.15: Diffusion coefficients of VWF C3 and the linker impurity 92
Figure 2.16: TrxA-His_6-VWFC3 (EK) construct sequence 93
Figure 2.17: Purification of the uncleaved TrxA-His_6-VWFC3 (EK) 94
Figure 2.18: Small-scale cleavage and purification of VWF C3 95
Figure 2.19: Comparison of VWF C3 (EK) ^{15}N-HSQC to VWF C3 (FXa) 97
Figure 2.20: VWF C3 14-day stability test 98
Figure 2.21: Purification of isotopically-labelled VWF C3 99
Figure 2.22: Backbone amide assignment of VWF C3 100
Figure 2.23: The assignment from K2429 to the EK cleavage site 101
Figure 3.1: Transition from a backbone $^{1}H_{N}{}^{15}N_{H}$ assignment to a $^{1}H_{\alpha}{}^{13}C_{\alpha}$ assignment 104
Figure 3.2: Assignment of the aliphatic ^{1}H and ^{13}C nuclei of I2437 using the HCCH-TOCSY experiment 106
Figure 3.3: Stereospecific assignment of the V2430 and L2485 side chain nuclei using fractional labelling 108
Figure 3.4: ^{13}C-HSQC assignment of the VWF C3 $^{13}C_{\alpha}$ region 109
Figure 3.5: TALOS+ secondary structure prediction and NOE assignment 111
Figure 3.6: Assignment of β-sheet NOEs 112
Figure 3.7: Chemical shift variations of cystine $^{13}C_{\beta}$ nuclei and VWF C3 backbone rigidity 114
Figure 3.8: NOE assignment between the $^{1}H_{\alpha}$-$^{1}H_{\beta}$ nuclei of C2451 and C2468 115
Figure 3.9: NOE assignment between $^{1}H_{\beta}$ nuclei of C2448 and C2490 116
Figure 3.10: NOE assignment indicating a C2477-C2494 disulphide bond 116
Figure 3.11: NOE assignment between the $^{1}H_{\beta}$ nuclei of C2473 and C2491 117
Figure 3.12: A NOE assignment indicating a C2431-C2453 disulphide bond 117
Figure 3.13: Arrangement of the VWF C3 disulphide bonds 118
Figure 3.14: Assignment of the side chain amide groups of glutamine, tryptophan and arginine 119
Figure 3.15: Assignment of the Y2483 aromatic nuclei 121
Figure 3.16: Tertiary structure NOE assignment between residues H2433, Y2438 and W2444 123
Figure 3.17: Tertiary structure NOE assignment 124
Figure 3.18: NOE assignment between R2478 and F2481 124
Figure 3.19: Assigned NOEs for VWF C3 residues 125

Figure 3.20: Residual dipolar coupling measurements using an in-phase/anti-phase (IPAP) experiment 126
Figure 3.21: High resolution structure of the VWF C3 domain 128
Figure 3.22: Accessibility of VWF C3 residues 130
Figure 3.23: Twenty lowest energy structures of VWF C3 131
Figure 3.24: Salt bridge between E2474 and R2478 132
Figure 3.25: Overlay of the SD1 of VWF C3 and the VWFC domain of CCN3 GF 133
Figure 3.26: Structural comparison between the full-length VWF C3 domain and CCN3 GF VWFC domain 134
Figure 3.27: Structural comparison between the full-length VWF C3 domain and VWF C4 domain 135
Figure 3.28: Overlay of the SD1 of the VWF C3 and C4 domains 136
Figure 3.29: Comparison of the twenty lowest energy structures of VWF C3 and C4 137
Figure 3.30: TrxA-His_6-VWFC4 construct sequence 138
Figure 3.31: Purification of ^{15}N-labelled VWF C4 139
Figure 3.32: ^{15}N-HSQC overlap of the VWF C3 and C4 domains 140
Figure 3.33: Purification of the C3L protein 142
Figure 3.34: Overlay of the C3NL and C3L ^{15}N-HSQCs 143
Figure 3.35: Backbone chemical shift changes between the C3NL and C3L proteins 144
Figure 4.1: Locations of the VWD-causing substitution mutations in the C3 domain 146
Figure 4.2: Purification of the VWF C3 C2431A/C2453A mutant 148
Figure 4.3: ^{15}N-HSQC of the C3 C2431A/C2453A mutant 149
Figure 4.4: ^{15}N-HSQC of the C3 C2451A/C2468A mutant 150
Figure 4.5: SD2 peak overlay in the SD1 disulphide bond mutants 151
Figure 4.6: Purification of the C2477A/C2494A C3 mutant 153
Figure 4.7: ^{15}N-HSQC of the C3 C2473A/C2491A and C2477A/C2494A mutants 154
Figure 4.8: Purification of the C2448A/C2490A mutant 156
Figure 4.9: ^{15}N-HSQC of the C3 C2448A/C2490A mutant 157
Figure 4.10: SD2 peak overlay in the SD1-SD2 disulphide bond mutant 158
Figure 6.1: Stejskal-Tanner equation to determine the diffusion coefficient of a selected peak 174
Figure 6.2: Calculation file for the CYANA structure calculation 179

List of Tables

Table 1.1: Classification of the VWD types and subtypes 59
Table 2.1: Ellman's Assay for the free thiol concentration of the expressed protein..... 84
Table 2.2: Experiments used for the backbone $^{1}H_{N}{}^{15}N_{H}$ assignment 86
Table 3.1: Experiments used in the side chain $^{1}H^{13}C$ chemical shift assignment..... 103
Table 3.2: Unassigned ^{1}H nuclei of VWF C3 110
Table 3.3: 3D NOESY experiments for the secondary and tertiary structural assignment 110
Table 3.4: Ellman's Assay for the free thiol concentration of the C3 protein..... 113
Table 6.1: Buffers used in the Ni-NTA and gel filtration purification..... 166
Table 6.2: PCR site-directed mutagenesis reaction setup..... 168
Table 6.3: PCR site-directed mutagenesis thermocycling conditions 168
Table 6.4: DpnI digest reaction setup 169
Table 6.5: M9 minimal media for expression of isotopically-labelled proteins..... 171
Table 6.6: Buffers and cysteine standards for an Ellman’s assay preparation..... 172
Table 6.7: NMR experiments for the backbone and side chain chemical shift assignment 173
Table 6.8: Buffers used in the purification of the C3 protein 175
Table 6.9: NMR experiments for the backbone and side chain chemical shift assignment 176
Table 6.10: NMR experiments for the NOE assignment 177
Table 6.11: Thermocycling conditions of the PCR site-directed mutagenesis of C3 cysteine residues 182

1. Introduction

Present day knowledge of blood and the cardiovascular system is extensive and improving year on year, however, ancient people knew nothing of the biology we know nowadays. Blood was only visible when weeping from a wound or during childbirth, becoming a symbol of life and death. This explains why blood became so heavily prominent in many religious traditions. The circulation of blood was first discovered in 1628 by English physician William Harvey. Along with disproving the theory that the body consumed blood, rather pumping 245 kg around the body per day, Harvey also showed that blood flows in two distinct loops, the pulmonary and systemic circulations[1]. In the years that followed, the first blood transfusions were attempted, the first successful transfusion being by Richard Lower in 1666 who performed the procedure on a dog[2]. It was not until the 19th century, however, that key components of blood coagulation, including platelets and tissue factor (TF), were discovered with the first description of coagulation presented by Paul Morawitz in 1905[3].

1.1. Haemostasis and the Coagulation Cascade

Haemostasis is defined by maintenance of blood fluidity throughout the vascular system by the cessation of blood loss, carefully designed to maintain laminar blood flow, prevent fatal haemorrhage and minimise the vascular system's exposure to foreign bodies. The term derives from Ancient Greek, "heme" meaning blood and "stasis" meaning halting. It is a vitally important physiological mechanism comprised of a series of intricately designed processes primarily involving the endothelium, platelets and coagulation cascade[4]. Haemostasis can be divided into primary, the formation of a clot by platelets acting as the first line of defence, and secondary, the repair of the vessel wall through the coagulation cascade. The coagulation cascade involves a series of enzymes which activate the proteins in the next step of the cascade, finally resulting in the polymerisation of fibrin and activation of platelets, leading to a blood clot[5] (Figure 1.1). It was first described using the Cascade model, presented in 1964, using a simple waterfall sequence to describe the basic mechanism of blood's ability to clot via a cascade of enzymatic activity[6, 7]. The Cascade model described the phosphatidylserine-containing cell surfaces as somewhere on which the procoagulant complexes were formed with coagulation factors being the primary controlling factor for haemostasis, however, the cell-based model, an updated theory presented in 2003, suggested otherwise. It proposed that these cell surfaces play a principal role in controlling haemostasis with the interaction of

clotting factors with specific cell surfaces occurring in three overlapping steps: initiation, amplification and propagation[8].

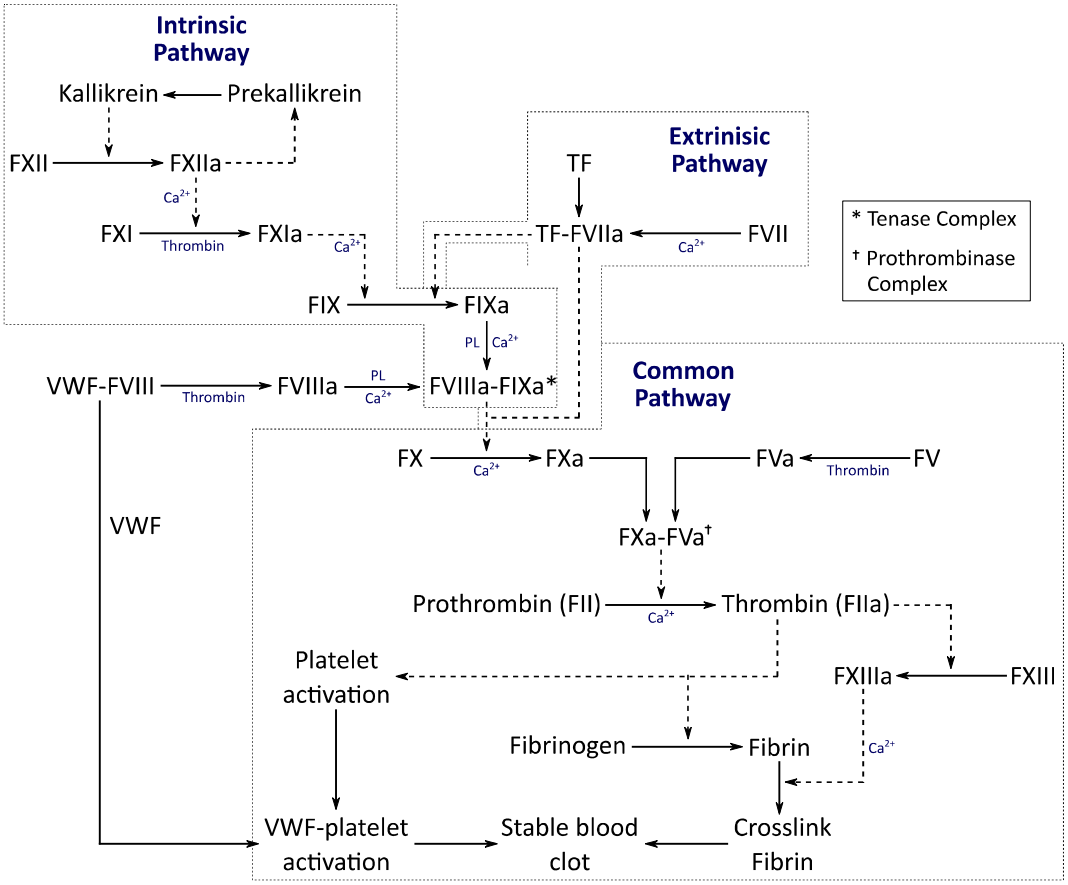


Figure 1.1: Overview of the coagulation cascade A summary of the coagulation cascade leading to the formation of a stable blood clot. Abbreviations used are: F – Factor, Ca^{2+} - calcium, TF – tissue factor, PL – phospholipid, VWF – von Willebrand factor. Symbols used indicate: * - Tenase Complex, † - Prothrombinase Complex. Solid arrows indicate a proteolytic conversion with dotted arrows indicating a reaction cofactor. Other cofactors in a reaction are shown in blue. The initiation of the cascade by the extrinsic and intrinsic pathways are indicated. The extrinsic, more dominant pathway is initiated as TF binds and activates FVII and subsequently acts as a cofactor in the activation of FX. The intrinsic pathway begins with the activation of FXII, increased by a positive feedback loop with kallikrein, and results in the activation of FXI, FIX and the formation of the Tenase Complex. The two pathways meet to activate FX, the start of the common pathway. FXa and FVa form the Prothrombinase Complex which converts prothrombin to thrombin. Subsequent activation of FXIII, platelet activation and formation of fibrin leads to a stable blood clot.

Initiation occurs via the extrinsic pathway on the TF-bearing cell, named such as under normal circumstances this resides outside of the vascular system. Upon contact with blood via injury or inflammation, the cells expose TF which subsequently binds and activates coagulation Factor VII (FVII), the beginning of the amplification step. The product, TF-activated FVII (FVIIa), cleaves FIX to FIXa and FX to FXa, the start of the common pathway where the two initiation pathways meet[8]. Whilst the extrinsic pathway is far more significant, the blood clotting cascade can also be triggered by the intrinsic pathway. Following a lesion, blood comes into

contact with an activation surface, known as the contact phase[9]. This causes the activation of FXII (to FXIIa), with subsequent activation of prekallikrein to kallikrein and further activation of FXII in a positive feedback loop[10]. This positive feedback loop is ~30x more efficient at activating FXII than autoactivation[11]. Activated FXII activates FXI which subsequently activates FIX, with FIXa forming a Tenase Complex with FVIIIa. This, along with TF-FVIIa, acts as a cofactor in the activation of FX[12]. Whilst in vitro experiments show the intrinsic pathway as being significant, humans lacking FXII do not experience excessive bleeding[13]. This may suggest that the contact pathway is more important for its production of the inflammatory bradykinin, produced by the cleavage of high molecular weight kininogen and an important vasodilator[5].

The common pathway begins when FX is activated. FXa binds to FVa, activated from FV by low levels of thrombin produced on the surface of TF-bearing cells, forming the FXa-FVa Prothrombinase Complex which converts prothrombin to thrombin. The increasing levels of thrombin propagates throughout the coagulation cascade, further activating FV, converting FVIII to FVIIIa and FXI to FXIa which acts as a cofactor in the activation of FIX. In parallel with this, the increased thrombin levels on TF-bearing cells further activates platelets, which have been concentrated on the site of injury via the action of von Willebrand factor (VWF). The intrinsic pathway controls the final step of the cell-based model which allows thrombin generation on the cell surface of activated platelets to propagate. Activated FVIIIa and FIXa form a tenase complex, which activates FX and forms the subsequent prothrombinase complex with FVa. Provided a significant increase in thrombin generation occurs through this complex, it results in the formation of a fibrin clot[8].

Through a number of coagulation inhibitors, a delicate balance is maintained between the pro-thrombotic pathway and the unaffected areas of the vascular system beyond the site of injury in order to prevent unnecessary clot propagation[14]. There are a number of natural inhibitors of clotting factors including antithrombin III which inhibits coagulation by neutralising the enzymatic activity of thrombin[15, 16]. Another inhibitor is the endogenous tissue factor pathway inhibitor (TFPI) which is primarily expressed in endothelial cells. This targets the initiation phase of coagulation, inactivating the majority of FXa bound to TF-FVIIa[14].

1.2. Von Willebrand Factor

VWF is a large multimeric glycoprotein found predominantly in plasma. Erik von Willebrand first recorded a hereditary bleeding disorder in 1924 whereby four sisters, three of whom had died, all presented a severe bleeding phenotype from birth. It was discovered in the mid-1950s that 'human plasma fraction 1-0' could correct deficiency of FVIII and prolonged bleeding time. Not until 1971 was the factor responsible identified and named VWF[17]. Since von Willebrand disease (VWD) was shown to be due to a lack of an essential blood clotting factor, VWF, the protein and disease have been extensively studied and our understanding has improved immeasurably[18]. VWF is responsible for platelet adhesion to the site of vessel injury under high shear conditions in primary haemostasis, the first line of defence against haemorrhage. Once bound to exposed collagen at a vessel lesion, circulating globular VWF unfolds and elongates into its long-chain conformation, becoming highly reactive[19]. This allows it to recruit platelets and mediate their adhesive interactions, forming a thrombus over said lesion before a coagulation cascade-mediated fibrin mesh forms[18, 20]. The second essential role of VWF is to act as a carrier for blood coagulation FVIII which, as seen previously, is essential in the formation of the Tenase Complex. Otherwise rapidly broken down when circulating freely, FVIII is released from VWF by thrombin activation at the site of injury, localising FVIIIa to where secondary haemostasis is required[21].

Synthesis of VWF occurs primarily in the endothelium[22], with this being the main source of plasma VWF[23]. Endothelial-derived VWF is either released constitutively into the plasma or stored in cytoplasmic storage vesicles, known as Weibel-Palade bodies (WPBs)[24]. Megakaryocytes, platelet precursors produced in the bone marrow, also produce VWF, with it being stored in α-granules of the subsequent platelets[25, 26]. Evidence suggests that endothelial cell-derived VWF is sufficient to support normal haemostasis, however, platelet-derived VWF also contributes to controlling bleeding[27]. Production of VWF is not consistent throughout the body, with VWF mRNA between 5 and 50 times more concentrated in the lung and brain than in the kidney or liver. Generally, VWF mRNA is more prominent in larger vessels versus microvessels and in venous endothelial cells versus arterial[28]. Exceptions to this include the pulmonary vein, where it was found to be absent, and the aorta and pulmonary artery, both of which showed high levels of stored VWF[29]. Distribution within the aorta appeared to be concentrated at a group of cells that were orientated along the longitudinal axis of the aorta, which may suggest that this distribution is an effect of laminar flow[30]. The reasons for the

variation of VWF distribution remain unclear, however, it may be due to a number of contributing factors such as flow forces, blood pressure and blood oxygenation[29].

1.3. VWF Genetics and translation product

The human VWF gene is located on the short arm on chromosome 12 at locus 12p13.31[31]. It spans 178 kb over 52 exons, the largest of which is the 1.4 kb exon 28 which encodes both the A1 and A2 domains[32]. Chromosome 22 also contains a VWF pseudogene, an imperfect copy of the functional gene[33] which contains a number of nonsense and missense mutations that prevent the expression of a functional protein. It spans 21 to 29 kb in length, corresponding to exons 22 to 34 of the VWF gene which encode the A1, A2 and A3 domains[34]. VWF repeat domains are dispersed throughout the human genome by duplication and exon shuffling with 46 A domains dispersed throughout 22 different human genes and A domains subsequently found in leukocyte adhesion receptors, collagen receptors and nonfibrillar collagens[34-36].

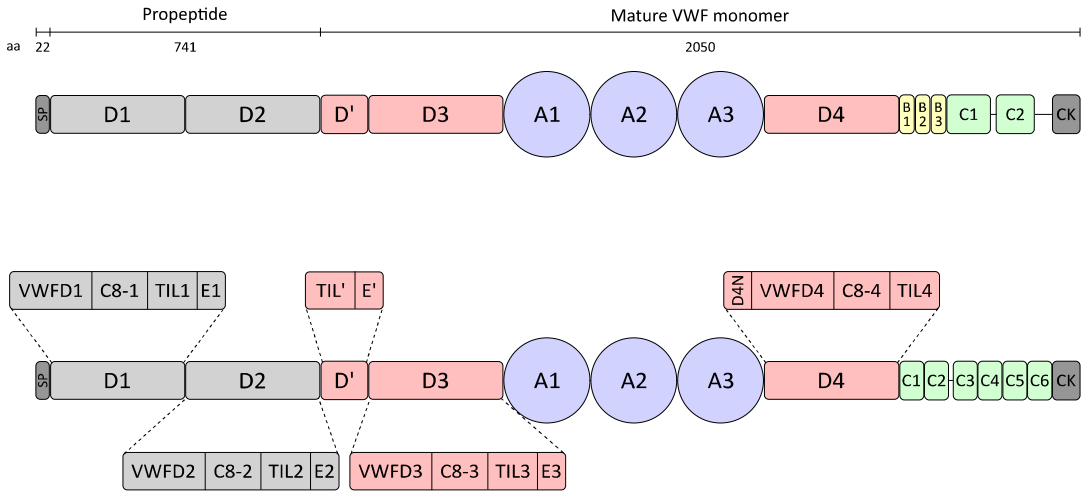


Figure 1.2: Monomeric VWF domain assignment Displayed is the original domain assignment, above, and the current assignment after analysis of sequence relationships, below[37]. In dark grey are the signal peptide (SP) at the N-terminus and cystine knot (CK) at the C-terminus. The D1-D2 propeptide is shown in light grey and the mature monomer coloured in red (D domains), blue (A domains), green (C domains) and dark grey (CK domain). The subdomains that make up each D domain are also indicated.

The complete VWF amino acid sequence was first published in 1986 and its repeating domain structure became instantly apparent[38]. Encoded by the 9 kb VWF mRNA, the expressed 2813 amino acid prepropolypeptide has a mass of ~350 kDa[39] (Figure 1.2) and is produced in the endothelial cells or megakaryocytes. The most abundant amino acid in VWF is cysteine, making up 8.3% of the protein, compared to just the 2.3% average in human proteins[40]. This highlights the importance of disulphide bonds in forming a stable VWF structure. The ProVWF domains were initially classed with between two and five repeats of four homologous domain types (A,

B, C and D), with a 741-residue propeptide, D1-D2, all preceded by a 22-residue signal peptide (SP)[41]. The three A domains are known to have a globular structure with a hydrophobic core, with the remaining domains appearing smaller and lacking a hydrophobic core, accounting for their richness in disulphide bonds[37]. At the C terminus is a cystine knot (CK) domain, homologous to the CK family of proteins which are often found forming disulphide bonds between monomers in dimerisation[42]. The VWF CK domain is a 90-residue domain which comprises 11 cysteines, particularly similar to those found in epithelial mucins and Norrie disease proteins which form C-terminal oligomers[43].

The domain reassignment of the VWF D domains was carried out with the principles that disulphide bonds in extracellular proteins are primarily intra-domain and that inter-domain sequences do not contain unexplained cysteines. Each D domain is subdivided into repeating VWF Type D (VWFD), C8, TIL and E modules, with the E module initially being touted as a domain type, however, it was soon classified as the C-terminal repeat within the D domain structure. Each of D1, D2 and D3 contain all four module types (D1 comprised of VWFD1, C8-1, TIL1 and E1 modules etc), whereas D' solely contains a TIL' and E' module and D4, whilst lacking an E segment, also contains a unique D4N segment at its N terminus. Each VWFD subdomain contains an even number of disulphide bonds except VWFD4 with its one odd cysteine bound to that in D4N[37]. The C8 domain, also found in crossveinless-2 (CV2) and mucins among others[44, 45], contains ten cysteines, with C8-3 and C8-4 containing an eleventh cysteine, forming an inter-dimer disulphide bond and inter-subdomain disulphide bond with TIL4, respectively[37]. VWFD, C8 and TIL domains show similarities to many domains found beyond VWF, with the structure of TIL domains well characterised prior to the annotation of VWF in this manner[46]. Each TIL module is followed by an E module which shows similarity to the first subdomain found in VWF Type C (VWFC) domains. The exception to this, however, is TIL4 with D4 lacking an E module though may be explained with TIL4 immediately proceeded by the C1 domain. The E repeats contain either four or six cysteine residues forming inter-subdomain disulphide bonds, as is seen in the first subdomain of VWFC domains[37].

Whilst initial domain assignment classed the C terminal region with a B1-B2-B3-C1-C2 arrangement, advancements in structural and bioinformatical analysis saw a reclassification of this region with 6 repeating C domains (C1-C6)[41]. Three domains in the C-terminal region were identified as VWFC domains, a protein domain found in various plasma proteins consisting of two tandem subdomains of 75-100 residues and 10 conserved cysteine residues[47]. This domain type is one of the most common to be found in extracellular proteins, accounting for 201 domains in 70 human proteins[48]. VWFC domains also contain specific cysteine motifs, CXXCXC

laying in the middle and CCXXC laying towards the C-terminal region[47]. The three VWFC domains identified in the C-terminal region of VWF, now annotated as C1, C3 and C5, were separated by three intervening regions, the middle of which contained the RGD motif, the binding site for $\alpha_{IIb}\beta_3$. These regions were similar to the VWFC domains in both length and sequence, and were therefore annotated at C2, C4 and C6. Outside of these domains, an intervening 26 residue region is also present between C2 and C3, rich in serine and threonine and containing 4 self-associated cysteine residues. The C domains, each of which contain between 8 and 13 cysteines, rely heavily upon disulphide bonds for their secondary structure rather than adopting a globular form as seen in the A domains. The majority are predicted to be intra-domain disulphide bonds, however, with an uneven number of cysteines in C1 and C2, a disulphide bond is predicted across the junction between the domains. Interactions at the C3-C4 and C5-C6 junctions are also likely, with the last 9 residues of C3 and C5 being identical, and the first 3 residues of C4 and C6 being almost identical[37]. With the high flexibility seen in VWFC domain structures, such as in collagen IIA (Col2A) determined using nuclear magnetic resonance (NMR) spectroscopy, the VWF C domains likely have a rope-like character, being highly flexible both within the domain and between the junctions[49].

1.4. Primary VWF functions and binding sites

A vast number of proteins are comprised of many tandem domain repeats, related units that have developed through evolution. These repeats have certain properties that characterise them but can vary in sequence and structure with certain sequences critical to their functionality[50]. The repeating domains often play a vital role in functionality, as is the case with VWF with multiple repeating domains, and subdomains in the case of the D domains, each of which contain their own unique binding sites (Figure 1.3). These binding sites allow VWF to carry out its primary functions as a bridging molecule between collagen and platelets in haemostasis[51] and acting as a carrier for blood clotting FVIII[52]. Alternative binding sites also allow the cleavage of VWF multimers by ADAMTS13 (**a** **d**isintegrin **a**nd **m**etalloproteinase with a **t**hrombo**s**pondin type 1 motif, member **13**)[53], the co-residence of P-selectin in WPBs and α-granules with its binding to VWF[54, 55] and the inhibition of VWF by heparin binding[56].

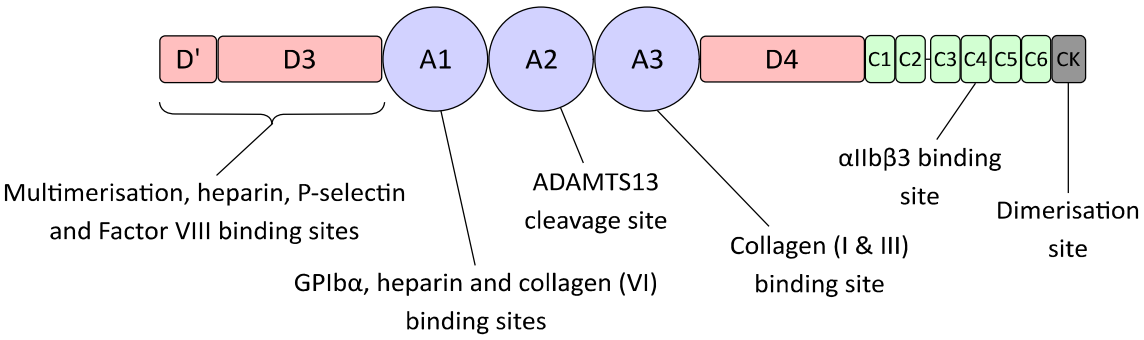


Figure 1.3: Binding sites of the mature VWF monomer Displayed are the binding sites found in different domains of the VWF monomer. The multimerisation site is found in the D'D3 region, along with the heparin, P-selectin and FVIII binding sites. A heparin binding site is also found in the A1 domain, alongside a GpIbα binding site for platelet binding and a collagen (VI) binding site. The cleavage site for ADAMTS13, the VWF cleavage enzyme, is found in the A2 domain with a collagen (I and III) binding site in the A3 domain. The $\alpha_{IIb}\beta_3$ binding site, responsible for platelet aggregation, is found in the C4 domain and the dimerisation site in the C-terminal CK domain.

1.4.1. D domains

With one of the main two functions of VWF being a carrier for coagulation FVIII, locating and understanding the binding site is of paramount importance. FVIII is comprised of a heterodimer in circulation consisting of a heavy chain (A1-a1-A2-a2-B) and a light chain (a3-A3-C1-C2)[57], non-covalently associated via the A1-A3 domains in a metal-ion dependent manner[58]. Activated FVIII forms one half of the Tenase Complex in the coagulation cascade, as shown in Figure 1.1, which in turn activates FX[59]. VWF performs a protective role when non-covalently bound to FVIII in circulation with genetic variants of VWF which impact FVIII binding leading to reduced levels of FVIII in circulation. This is classified as Type 2N VWD[60]. Prior to the discovery of VWF, reduced levels of circulating FVIII classified the patient as having haemophilia A, however, it was soon discovered that this was due to insufficient binding of FVIII to VWF and led to a rapid FVIII breakdown[61]. The D' domain has been shown to bind and stabilise FVIII and its NMR structure confirmed its splitting into two distinct subdomains, the trypsin-inhibitor-like' (TIL') and E' domains. TIL' is comprised of two short antiparallel β-sheets, β1:β2 and β3:β4, which form a scaffold for the long β1-to-β2 loop. The E' domain adds to this scaffold structure for the long β1-to-β2 loop, containing two antiparallel β-sheets, one double stranded and one triple-stranded. Both subdomains are heavily disulphide bonded with TIL' containing five intra-subdomain disulphide bonds and E' containing three. Despite the lack of disulphide bonding between the subdomains, the TIL' and E' domains tumble with limited inter-domain motion with a high degree of flexibility only seen in the TIL' β1-to-β2 loop[62].

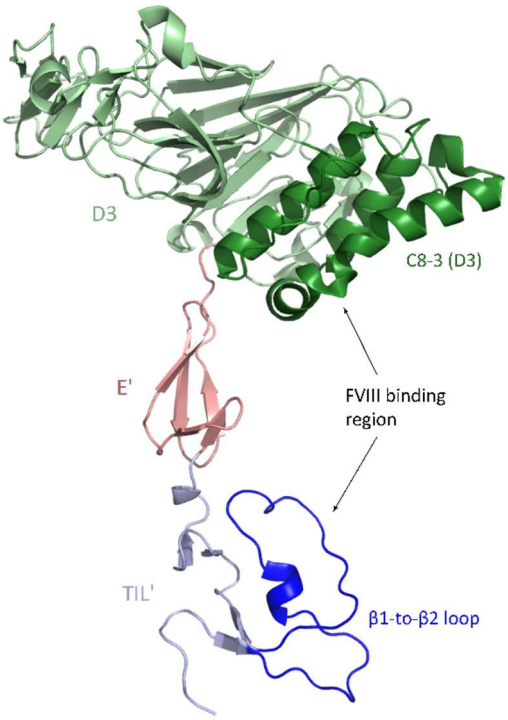


Figure 1.4: Structure of D'D3 region The structure of the D'D3 region is displayed, with the TIL' subdomain in blue, the E' subdomain in red and the D3 domain in green. Highlighted in a darker shade is the FVIII binding region, comprised of the β1-to-β2 loop of TIL' (dark blue) and C8-3 subdomain of D3 (dark green)[62, 63].

A crystal structure of the D'D3 domains was also developed, with the D' domain of the crystal structure matching that of the NMR structure[62, 63] (Figure 1.4). The D3 domain was shown to have its larger VWFD3 module at the centre and the smaller C8-3, TIL3 and E3 modules wrapped around it to form a wedge-shaped assembly. The central VWFD3 module is formed by 12 β-sheets and one α-helix, leavings strands β1, β2, β3, β9, β10, β11 and β12 forming the face for the other modules to wrap around, with β4, β5, β6, β7, β8 and α1 solvent exposed. The TIL' and E' modules sit end to end at an angle of 60° from the rectangular face of the wedge-shaped D3 assembly[63]. This is consistent with the horn-shaped protrusion from the surface of D3 observed in electron microscopy (EM) studies[64, 65]. Inter-domain and inter-subdomain flexibility is limited in the D'D3 assembly, firstly with a Cys-Pro-Cys motif at the TIL'-E' junction, with both cysteine residues separately disulphide bonded. Even less flexibility is seen between E' and VWFD3, the junction stabilised by a disulphide bond and van der Waals contact of V862 of E' with Q989 and F999 of VWFD3. This is also seen in the interactions of VWFD3 with the other D3 modules, particularly the strongly hydrophobic interaction of VWFD3 with C8-3 and TIL3, supporting the hypothesis that D assemblies are created by the highly specific associations between modules[63].

Binding of VWF to FVIII has been shown to occur in the TIL' module of the D' domain, specifically to the flexible β1-to-β2 loop region of the domain and supported by the β-sheets of the TIL' and E' subdomains[62, 66]. It was suggested that this positively charged dynamic region binds to the negatively charged FVIII a3 domain[62] with a ~8x greater binding affinity to VWF than the FVIII C2 domain[67]. The major FVIII binding region of VWF being the flexible region of TIL' was supported by EM studies, however, determined that the role of the acidic a3 domain in binding affinity needed further investigation[64]. Genetic variants leading to Type 2N VWD are not limited to the TIL' FVIII binding region itself as the variants identified span across a 486 amino acid region. Although, deletion of the E3 subdomain does not appear to have an impact on FVIII binding, many mutations across VWFD3 and C8-3 cause a conformational change in the D3 domain which correlates with reduced FVIII binding affinity[68]. Type 2N mutations in C8-3 include Q1053H and E1078K which decrease FVIII binding affinity tenfold and more than tenfold, respectively[69, 70].

EM studies along with these binding affinity observations suggest that there are two non-contiguous binding sites between the FVIII C1 domain and VWF[64, 65]. The crystal structure of the D'D3 region further suggested the secondary role of VWF D3 in FVIII binding. Two binding sites separated by 5 nm were identified with the FVIII C1 domain interacting with TIL' and a secondary site in C8-3 encompassing much of the face of the helical bundle containing the α1 and α3 helices. This is consistent with the fact that TIL' is located closer to C8-3 than any other D3 module[63]. Other mutations that cause VWD Type 2N include D879N in the Ca^{2+} binding site of the D3 domain (along with VWD Type 2A which will be discussed in Chapter 1.7.1), and R854Q in the E' domain[71]. With TIL' residing closer to C8-3 than any D3 module, these mutations can be explained as altering the TIL'-E' and E'-VWFD3 orientation and, hence, the orientation of the binding sites in TIL' and C8-3[63].

Along with minimising FVIII breakdown in circulation, VWF also plays a role in limiting the action of FVIII, preventing binding of FVIII to activated FIX, as inactivated FVIII has a binding affinity to VWF 100-fold more than that of FIXa[72]. Activation of FVIII is understood to be a multistep process, where cleavage of FVIII from its VWF binding site occurs in multiple steps[73]. FVIII activation is primarily induced by thrombin[74] with FXa also shown to be an efficient activator[75], though FXa-induced FVIII has been reported as displaying reduced activity compared to thrombin-activated FVIII[76]. This can be explained as FXa-induced activation is slowed when FVIII is associated with VWF whilst there is no effect of VWF on the rate of thrombin-induced activation[77]. Both thrombin and FXa cleave in both the light and heavy chain of FVIII, with thrombin cleaving at R1689 in the light chain and R372 and R740 in the heavy

chain. Cleavage by FXa also occurs at these 3 sites, along with R1721 in the light chain and R336 in the heavy chain[75]. This results in the removal of the B domain and the dissociation of FVIIIa from VWF[78] in a 1600-fold decrease in FVIII affinity for VWF[67]. This in turn favours the binding of FIXa to FVIIIa and the subsequent formation of the Tenase complex[57]. Recent evidence has also suggested that VWF binding to myosin in the setting of trauma may indirectly accelerate thrombin generation by providing a surface for VWF binding and facilitating the delivery of FVIII to the site of injury[79].

1.4.2. A domains

The VWF A domains are more globular than other VWF domains, characterised by far fewer disulphide bonds and a hydrophobic core[37]. They house the binding sites for VWF to carry out its second primary function, binding to both collagen and platelets in order to recruit platelets to the site of injury, known as primary haemostasis. Collagen binding occurs in both A1 and A3 domains, however, A3 domain collagen binding to types I and III has been shown to be the principal binding site[80]. The GpIbα platelet binding site is located in the A1 domain, with the adhesive properties tightly regulated in order to prevent sporadic binding circulating platelets[81]. The importance of collagen binding was shown when blocking of the VWF-collagen interaction resulted in the abolition of platelet thrombus formation[82]. VWF binding to subendothelial collagen, exposed by vascular injury, results in a conformational change which exposes the GpIbα binding site for platelet binding to occur[81]. VWF type A domains are found in a vast array of proteins, including αLβ2 and αMβ2, commonly referred to as the I domains in those proteins[83, 84]. The proteins where these I domains are formed are involved in a variety of biological functions, including cell-cell interactions and ligand-receptor binding[85]. These are characterised by a central hydrophobic parallel β-sheet, flanked by an amphipathic α-helix on either side, known as a dinucleotide binding fold[83]. At the top of the β-sheet is a metal ion-dependent adhesion site (MIDAS) motif which engages a glutamate residue of collagen. This motif is located in a groove where the I domains carry out their binding function, making it a critical feature[86, 87].

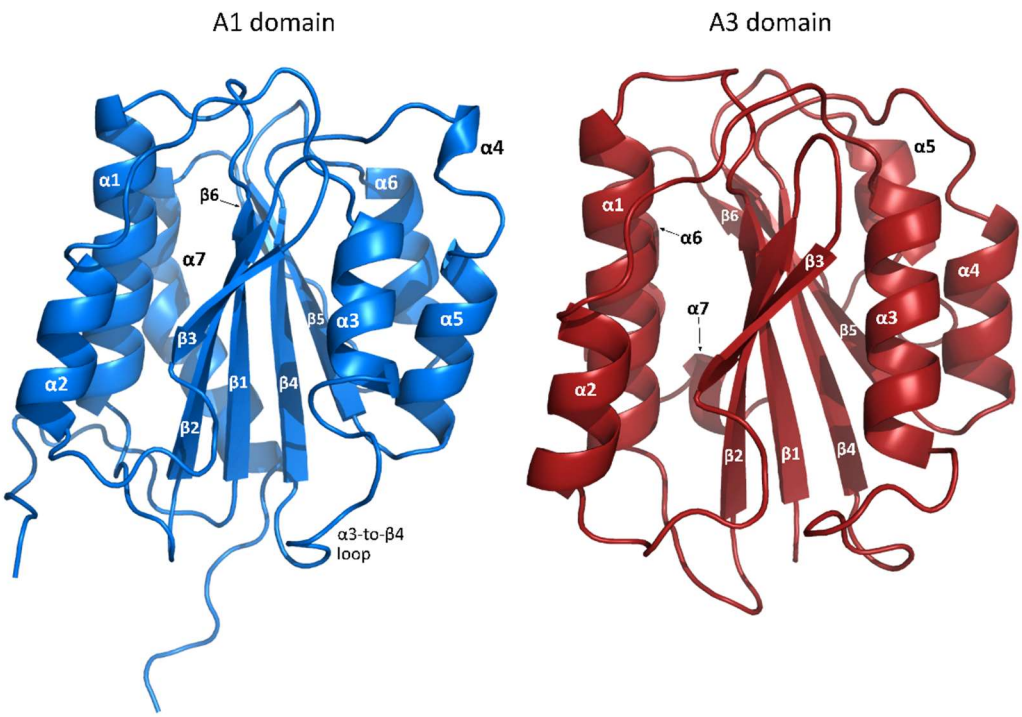


Figure 1.5: Structure of VWF A1 and A3 domains Displayed in blue is the crystal structure of the VWF A1 domain[88] and in red is the A3 domain[89]. Both are comprised of a central β-sheet region surrounded by seven amphipathic α-helices. The number of each β-sheet and α-helix is indicated.

The crystal structure of the VWF A3 domain (Figure 1.5) shows that whilst organised in a similar manner with six-stranded central hydrophobic parallel β-sheet flanked by seven α-helices, there are some differences between itself and the I domains. Whilst I domains contain a groove at the MIDAS motif, crystal structures of the VWF A3 domain show a smooth surface over this region and the rest of the domain[89]. Despite containing a MIDAS motif, the motif in the VWF A3 domain has been shown to be conformationally incompatible for metal ion binding and mutations in this region do not reduce collagen binding, meaning collagen binding occurs independent of this motif[81, 90]. The collagen binding site in the VWF A3 domain was located at the front, lower face, with the front face comprised of the α2 and α3 helices and β3 strand and the bottom face containing several loops and housing both the N and C termini. This site is noticeably distant from the binding site in the I domains at the top around the MIDAS motif. The H1023A mutation, located in loop α3-β4 at the edge between the front and bottom face, completely abolished collagen binding, stressing the importance of this site. Other residues also implicated in collagen binding were R963 and R1016, with mutations to alanine at these positions resulting in reduced binding[91]. Further studies of other mutants narrowed down the binding region to strand β3 and loop α3-β4, giving a rather hydrophobic binding site in contrast to the hydrophilic groove in the I domains and interacting with collagen sequences containing both positively charged and hydrophobic residues[92].

Residues in the lower half of the binding site, D797, S1020 and H1023, were deemed critical for collagen binding, whilst those in the top half, I975, T977, V997 and E1001, contributed to collagen binding but were deemed non-essential[92]. NMR studies of the VWF A3 domain determined that collagen binds in a transverse manner across the front lower region of the domain[93]. It was considered whether, despite the MIDAS motif not playing an active role in collagen binding, F939 located close to this motif could be relevant for the putative interaction between the A3 and A1 domains to influence platelet GpIbα binding[91]. Collagen binding, however, did not appear to affect the affinity of the VWF A1 domain to platelet receptor GpIbα[92]. The positively charged residues in the VWF A3 collagen binding region interact with negatively charged residues on collagen itself. The lack of a binding groove, as in the I domains, prevents a large van der Waals surface interaction between the two surfaces, hence why a single domain has such a low affinity for collagen binding and why multimeric VWF is so important for a tight association with collagen[89, 94].

Collagen is the most abundant peptide in the human body, comprising approximately one-third of the total protein[95]. There have been 28 different types of collagen discovered, comprised of at least 46 different polypeptide chains, providing tensile strength to various tissues and being involved in cell adhesion, migration and maturation[96, 97]. It is comprised of three peptide α-chains wound around each other in a right-handed triple helix. with collagen type I formed by two α1 chains and one α2 chain and collagen type III formed by three α1 chains[97, 98]. VWF binds to four types of collagen, VWF A3 domain to types I and III in the tunica media and tunica adventitia and VWF A1 domain to type III and VI in the subendothelium[99]. The collagen sequence identified as the major, high-affinity binding site for VWF is RGQOGVMGF (O is hydroxyproline), a sequence which occurs in collagen type II, not found in the subendothelium, and type III at position 572-580 on the α1 chain. Whilst not found in collagen type I, a similar sequence in the α1 chain with a single O to A substitution is found, RGQ**A**GVMGF, at position 574-582. Whilst mutation studies found that this sequence did not bind to VWF, the α1 chain's α2 compatriot aligns alongside it, with an O residue in the α2 chain being coupled with the single A at position 577 in the α1 chain. This is located in the sequence RGEOGNIGF at positions 486-494, suggesting that both collagen types I and III bind VWF in the same manner[97, 100].

Whilst the interaction of plasma VWF to subendothelial collagen has been shown to be mainly mediated by the A3 domain[101], binding to collagen type VI occurs in the A1 domain[98]. Recent evidence also shows that VWF binds to collagen type IV in the A1 domain with an R1399H mutation affecting collagen type IV binding and resulting in increased bleeding in a murine

model[102, 103]. This A1 collagen binding is particularly important for patients with VWD mutations in the A3 collagen binding region, as it can assume the role of VWF A3 in recruiting platelets to the site of injury[104]. Studies of the isolated VWF A1 domain shows that it also binds to collagen types I and III, however, blocking of the A1 domain by an antibody in multimeric VWF had no effect on the binding to collagen. This can likely be explained as the cryptic site is exposed in an isolated A1 domain, however, in full-length VWF this site is shielded by either the A2 or A3 domain, or both. The A1 domain may, therefore, play a role enhancing VWF binding to collagen under high levels of shear stress[105]. The crystal structure of the VWF A1 domain shows it to be, as expected, similar to the VWF A3 domain and integrin I domain (Figure 1.5). It is comprised of a hydrophobic six-stranded parallel β-sheet, flanked on two sides by seven amphipathic helices. Similar to VWF A3, it appears to lack the metal ion binding capability of the MIDAS motif with the homologous loops not suitable for metal ion coordination[88]. A serine residue in the MIDAS motif is replaced by an arginine, R524, which may play a role in stabilising the buried D520, playing a role analogous to the metal ion[81].

Binding of VWF to collagen is the first step in recruiting platelets to the site of injury in primary haemostasis, with the second step being its binding to GpIbα on the surface of platelets. VWF binding to platelets requires a high shear rate generated by rapidly flowing blood with platelet adhesion at lower flow rates occurring by other interactions, such as between collagen and GpIa-IIa (integrin $\alpha_2\beta_1$) and between fibrinogen and GpIIb-IIIa (integrin $\alpha_{IIb}\beta_3$)[106-108]. This binding can also be induced by mutations causing greater VWF-GpIbα binding affinity (Type 2B VWD), desialation (removal of sialic acids – discussed in Chapter 1.5.1) and treatment with antibiotic ristocetin or the snake venom protein botrocetin. GpIb-V-IX is a complex on the surface of platelets and is comprised of four transmembrane subunits, GpIbα, GpIbβ, GpIX and GpV, with the binding site for the VWF A1 domain localised to the 290 N-terminal residues of GpIbα[109]. The structure of GpIbα consists of an N-terminal β-hairpin delimited by a conserved disulphide bond between Cys4 and Cys17, eight leucine-rich repeats, a disulphide-bonded loop and an anionic C-terminal region[110].

The crystal structure of the VWFA1-GpIbα complex concluded that binding occurs via two sites, the most extensive site being at the top of the A1 domain, where helix α3, loop α3-to-β4 and strand β4 interact with leucine-rich repeats 5 to 8 and the C-terminal flank of GpIbα[111, 112]. The secondary, smaller interaction occurs at the bottom face of the A1 domain, with loops α1-to-β2, β3-α2 and α3-β4 interacting with the N-terminal β-hairpin and first leucine-rich repeat[111], with both of these occurring by long-range electrostatic interactions. Initial studies suggested that GpIbα bound to the VWF A1 domain via the negatively charged binding

surfaces of the leucine-rich repeat concave face which forms a flexible loop at residues 227-241 and the anionic region[110], however, later studies showed deletion of the anionic region did not affect VWF A1 binding. This flexible loop, also termed a β-switch, forms a 16 residue β-hairpin from residues 227-242 and forms a continuous β-sheet with VWF A1 by aligning with its β3 strand[111]. Binding of VWF to GpIbα only occurs at a shear rate greater than 500-800 s^{-1}, with the interaction between the two ligands requiring a sufficiently fast on-rate to initiate platelet adhesion[99]. Once bound to collagen at the vessel wall where shear rates are highest, the hydrodynamic forces elongate the VWF multimer. This causes a shear-induced conformational change which elongates the VWF multimer and exposes the GpIbα binding site in the VWF A1 domain[113]. The majority of these studies were carried out prior to the VWF domain reassignment by Zhou et al[37], using truncated forms of the A domains. These forms do not include what is, following the domain reassignment, the N-terminal region of VWF A1, residues which have since been shown to suppress the VWF A1-GpIbα catch bond, decreasing pause times and increasing translocation velocities[114].

1.4.3. C domains

VWF plays a critical role in not only recruiting platelets to the site of vascular injury, but also in the progression of thrombus formation by its interaction with specific platelet surface receptors, namely the megakaryocyte and platelet-specific integrin $\alpha_{IIb}\beta_3$[115]. Integrins are heterodimeric glycoprotein complexes consisting of a non-covalently bound α and β subunit, comprised of a long N-terminal extracellular ectodomain, a transmembrane spanning helix and a short C-terminal cytoplasmic domain[116]. The GpIIb/IIIa receptor (integrin $\alpha_{IIb}\beta_3$) is one of the most abundant platelet surface receptors, representing ~15% of total surface platelet proteins, around 80,000 surface copies with additional storage pools in platelet α-granules[117, 118] (see Chapter 1.5.5). In circulation, $\alpha_{IIb}\beta_3$ of inactivated platelets remain in a low-affinity binding state in a closed conformation[119]. Once platelets are activated by VWF binding to GpIbα, an inside-out signalling mechanism is induced which transmits a signal originating in the cell cytoplasm to the external receptor ligand-binding domain[118, 120]. This is triggered by a number of pathways which converge on the activation of the small GTPase and Rap1, and the recruitment of proteins talin and kindlin[116]. These proteins bind to the intracellular C-terminal tail of the β3 subunit, promoting domain separation and unfolding of the extracellular ectodomain[121]. This conformational change in the binding receptor allows it to bind to a high-affinity binding ligand, leading to the formation of a platelet plug[122].

A number of ligands bind $\alpha_{IIb}\beta_3$ including fibrinogen, fibrin, fibronectin and VWF, each of which contain a common Arg-Gly-Asp (RGD) binding motif[123]. The one exception is fibrinogen which

contains a KQAGDV sequence on the C-terminus of the γ-chain[116]. The $\alpha_{IIb}\beta_3$ signalling mechanism is bidirectional, with an outside-in signal pathway being used following ligand binding, with activated $\alpha_{IIb}\beta_3$ acting as a signal transducer to the cell interior. This mechanism couples to a number of intracellular effectors, ultimately leading to platelet spreading, stable thrombus formation and clot retraction[122]. Platelet spreading is characterised by a rearrangement of the cortical actin cytoskeleton, cell rounding and attachment to an immobilised surface. This outside-in signalling mechanism results in an increased surface area and strengthened contact with the surface and neighbouring platelets as they form lamellipodia[124].

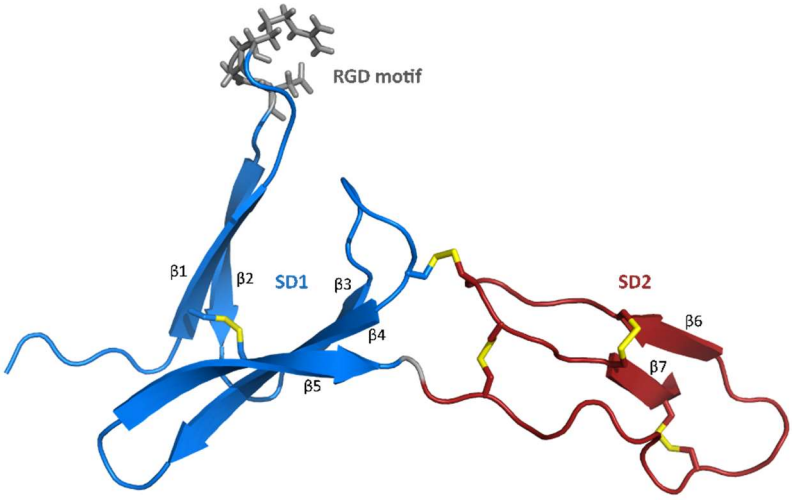


Figure 1.6: Structure of the VWF C4 domain Displayed is the NMR structure of the VWF C4 domain, with SD1 highlighted in blue and SD2 highlighted in red. The five disulphide bonds are highlighted in yellow with the RGD motif which binds to $\alpha_{IIb}\beta_3$ shown in grey[125].

The C domains of VWF provide integral support for platelet capture, heavily contributing to the flexibility and length that is vital for the transition between the globular and elongated states[126]. VWF is able to support platelet aggregation by binding to GpIbα to activate platelets and $\alpha_{IIb}\beta_3$ to initiate platelet aggregation[127, 128]. Binding to $\alpha_{IIb}\beta_3$ occurs at the RGD motif located in the C4 domain, the structure of which has been solved by NMR spectroscopy[125] (Figure 1.6). The overall fold of the VWF C4 domain comprises two separate subdomains – SD1 and SD2. SD1, which houses the $\alpha_{IIb}\beta_3$ binding site, is comprised of two antiparallel β-sheets, one double-stranded (β1 and β2) and one triple-stranded (β3, β4 and β5). The RGD motif is located at residues 2507-2509 at the tip of the 10-residue β1-to-β2 hairpin loop. Relaxation data and conformational heterogeneity of the 20 lowest energy structures shows the flexibility of the loop region, increasing the likelihood of an $\alpha_{IIb}\beta_3$ interaction. Disulphide bonds are an important factor in the VWF C4 structure, with one disulphide bond between the two β-sheets

in SD1 which is thought to be a fold determinant of the subdomain. Two further fold determinants are present within SD2 with a third disulphide bond thought to be unique to VWF C4, known as the "C4 bridge", connecting the two β-sheets of SD2. The final disulphide bond connects the two subdomains and is thought to determine the arrangement between SD1 and SD2. SD2 was also shown to have higher global mobility than SD1 with a hinge residue at position V2547, with flexibility between the two subdomains also highly sensitive to temperature[125].

VWFC domains are highly represented in matrix proteins which bind to bone morphogenetic proteins (BMPs) and regulate growth factor (GF) responses[47]. Whilst VWF C4 is termed a VWFC-like domain, the structure of a number VWFC domains of other proteins have been determined, including VWFC domains of Col2A, Cell Communication Network Factor 3 GF (CCN3 GF) and CV2[45, 47, 49, 129]. Despite VWF C4 presenting a high level of similarity to VWFC domains there are a number of differences, particularly in the cysteine and disulphide bond arrangement. The VWFC domain disulphide bond arrangement can be summarised, labelling residues Cys_1-Cys_{10}. The disulphide Cys_1-Cys_4 connects the N-terminal β1 strand to the β4 strand of the triple-stranded β-sheet. The β4 and β5 strands of the triple-stranded β-sheet are connected by Cys_3-Cys_5, the disulphide not found in VWF C4. The two disulphides in SD2, Cys_6-Cys_9 and Cys_7-Cys_{10}, connect strand β7 to the long sequence stretch which lacks a secondary structure. Finally, the two subdomains are connected by Cys_2-Cys_8[47]. VWFC domains typically contain two cysteine motifs, CXXCXC and CCXXC, towards the middle and C-terminal end of the domain, respectively. VWF C4 contains the CXXCXC motif at positions 2528-2533, however, is missing the middle cysteine at position 2531 due to its lack of a second disulphide bond in SD1[37, 49, 125]. Whilst VWF C4 contains a CCXXC motif towards its C-terminus at positions 2570-2574, the cysteine residue at position 2574 is involved in the unique "C4 bridge" that connects the two strands of the β-sheet, with a final cysteine residue at position 2576 involved in a similar disulphide bond to that in VWFC domains. In terms of those VWF C4 cysteine residues involved in the equivalent disulphide bonds to those in VWFC domains, the motif is actually summarised as CCXX**X**XC, with the **X** indicating the cysteine involved in the "C4 bridge"[125]. Considering these differences, when analysing the structure of the VWFC domains of VWF, C1, C3 and C5, it may be more useful to compare to the VWFC domains of Col2A, CCN3 GF and CV2, rather than VWF C4.

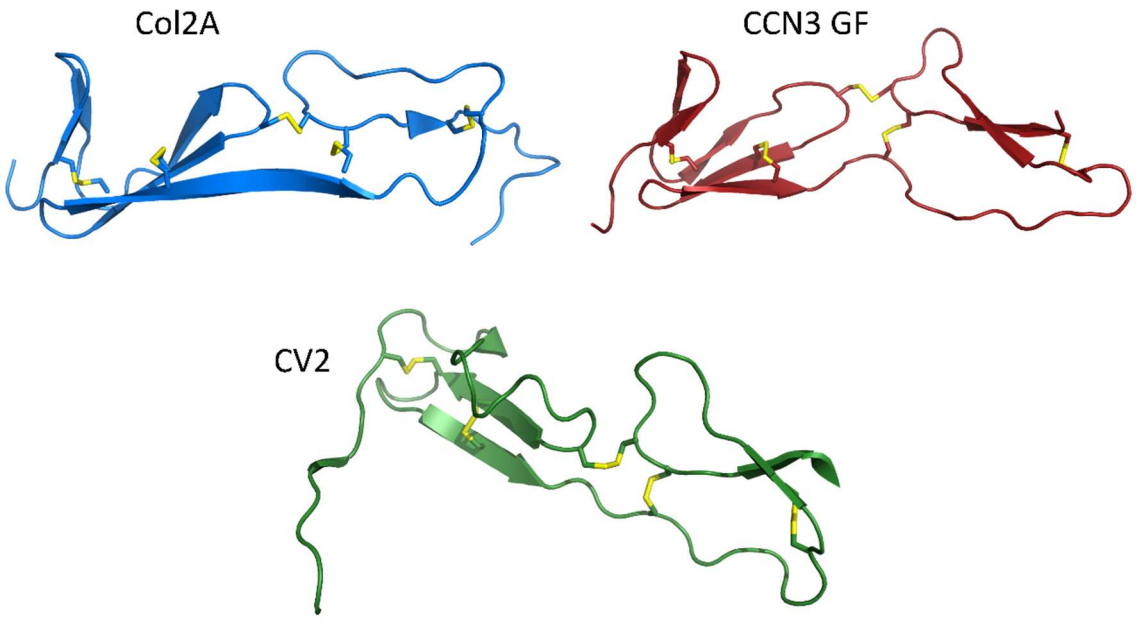


Figure 1.7: Structures of VWFC domains of Col2A, CCN3 GF and CV2 Displayed are the VWFC domains of Col2A (blue), CCN3 GF (red) and CV2 (green). Disulphide bonds are shown with S_γ of cysteine residues highlighted in yellow.

In terms of the structure of VWFC domains, accounting for the differences outlined above, they are by and large similar to VWF C4. The VWFC domain of Col2A was the first VWFC domain structure to be determined (Figure 1.7). Solved by NMR spectroscopy, it exhibits a two-subdomain structure with a β-sheet arrangement similar to that of VWF C4, with two intra-subdomain disulphide bonds in each subdomain and an inter-subdomain disulphide bond connecting them[49]. Relaxation data has shown the SD1 structure to be inflexible, largely dominated by β-sheets and stabilising disulphide bonds between them. The N-terminal β-hairpin structure is stabilised to the remainder of SD1 by a Cys_1-Cys_4 disulphide and the aromatic interaction between Tyr41 and Trp47. Interestingly, SD2 adopts a more irregular structure than that of SD1 with relaxation data confirming that SD2 is more flexible[47]. This is also observed in VWF C4, indicating that greater SD2 flexibility versus SD1 is a hallmark of the VWFC domains[125]. The structure of the CCN3 GF VWFC domain is consistent with that of Col2A with regards to the β-sheet and disulphide bond arrangement (Figure 1.7). A Cys_1-Cys_4 disulphide once again stabilises the N-terminal β-hairpin structure to the remainder of SD1 and a similar aromatic interaction between Y111 and F117. The CCN3 GF VWFC domain lacks a proline residue prior to Cys_2, a residue which is observed in the VWFC domains of Col2A and CV2. This causes a shift in position of the Cys_2-Cys_8 and subsequently Cys_6-Cys_9 bonds, however, this proline residue is only found in 25% of human VWFC domains, suggesting the disulphide positioning in CCN3 GF VWFC is likely more common. Clear differences arise between VWFC domains when looking at the orientation of SD2. The angles between the two

subdomains vary by 50° in the three domains discussed here, explained by the flexibility of the linker between the subdomains and the lack of the proline residue in CCN3 GF VWFC causing a shift in disulphide arrangement[47].

The root mean square deviation (RMSD) of a protein is used to measure the quality of a bundle of structures. That calculated of the CCN3 GF VWFC domain showed that the lowest energy structures were extremely similar with an average RMSD of 0.265 Å across the four lowest energy structures. In the case of the Col2A VWFC domain, however, incorporation of SD2 adversely affected the RMSD value. Whilst SD1 displayed a RMSD of between 0.9-1.3 Å for the four lowest energy structures, a much poorer alignment was seen when incorporating SD2, with an average RMSD of ~3.5 Å. This can be explained with the fact that Col2A VWFC is largely devoid of any secondary structure[47].

The VWFC1 domain of CV2 presents clear differences to other VWFC domains (Figure 1.7). Whilst not showing the same degree of flexibility between the two subdomains, the VWFC1 domain of CV2 shows a high degree of flexibility at its N-terminus. The CV2 VWFC1 domain is smaller than other VWFC domains being only 66 residues in length, however, it also contains an eight-residue N-terminal Clip segment[45]. This Clip segment comes at the expense of the N-terminal β-sheet of SD1 leaving only one triple-stranded antiparallel β-sheet (strands β1, β2 and β3) in SD1 and Cys_1 located on its N-terminal Clip rather than on the first β-sheet strand. CV2 VWFC1 also contains a bulged loop between strands β1 and β2 rather than being a regular β-sheet with a short hairpin as in the Col2A or CCN3 GF VWFC domains. These unique properties have been shown to be essential for its function, with the Clip segment and bulged loop found to be the two major BMP-2 binding epitopes, binding in a "paperclip" fashion[45, 129]. Recent evidence concluded that BMP-2 binding can occur in the triple-stranded antiparallel β-sheet region of Col2A VWFC1, albeit with far lower affinity than CV2, and not in CCN3 VWFC, highlighting the importance of the CV2 Clip segment and bulged loop in BMP-2 binding[47].

The secondary roles of VWF are discussed in Chapter 1.6.

1.5. Life Cycle of VWF

1.5.1. Dimerisation in the endoplasmic reticulum

The prepro-VWF molecule is translocated to the endoplasmic reticulum (ER) following translation, where the SP is cleaved to produce proVWF. Being a heavily disulphide-bonded protein[130], disulphide bonds within VWF multimers form in the ER, in the presence of thiol isomerases such as protein disulphide isomerase (PDI) and ERp57[131]. The ER exhibits a redox potential, suitable for disulphide bond formation, and chaperones which control disulphide bond reshuffling, in contrast to the Golgi which lacks known chaperones and is relatively acidic[132]. The ER also contains quality control mechanisms to ensure solely properly folded proteins are transported to the Golgi[131, 133]. During maturation, VWF undergoes N-linked glycosylation, the attachment of an oligosaccharide to the amide nitrogen of asparagine and is catalysed by the enzyme oligosaccharyltransferase. Twelve consensus sequons N-X-S/T in the mature VWF monomer were identified as being glycosylated (N857, N1231, N1515, N1574, N2223, N2290, N2357, N2400, N2546, N2585, N2635 and N2790)[38] with an additional rare sequon, N-X-C (N1147) later being identified as occupying N-glycans[134] (Figure 1.10). The fact that these glycosylation sites are located within, or around, the VWF functional sites makes it unsurprising that these structures are essential for proper synthesis or function of VWF, with initial glycosylation being essential for VWF multimerisation[135]. Glycans expressed at different positions within the monomer vary greatly in size, with N1515 and N1574 presenting larger, more complicated glycan structures, and N857 and N1147 presenting much smaller structures. Through mapping of the VWF N-glycome, ~300 structures have been identified revealing an immense variation in these glycan structures[134, 136, 137].

The most common observed glycan structures were monosialylated and disialylated biantennary chains, with over 90% of glycan structures capped by sialic acid structures, an acidic sugar with a nine-carbon backbone[134]. Sialic acid residues act to protect from post-translational VWF glycosylation[138]. The majority of this N-linked capping is present in an α2-6 linkage[134] (Figure 1.8). Whilst sialic acid capping is accepted to be essential in protecting VWF from proteolytic function, the removal of N-terminal sialic acid has also been shown to make VWF more resistant to ADAMTS13 cleavage, the metalloproteinase that regulates VWF function by breaking down large multimers through cleavage in the A2 domain. Through removal of specific linkages (α2-3,6,8,9) individually suggests that the most common linkage, α2-6, is the most critical for proteolysis by ADAMTS13[139]. Site-directed mutagenesis studies have identified the N-linked glycans at N1574 as particularly important due to its positioning in

the A2 domain close to the cleavage site at Y1605-M1606[140, 141]. Sialylation has also been shown to be critical in modulating VWF clearance, with removal of sialic acid residues significantly reducing VWF half-life[142].

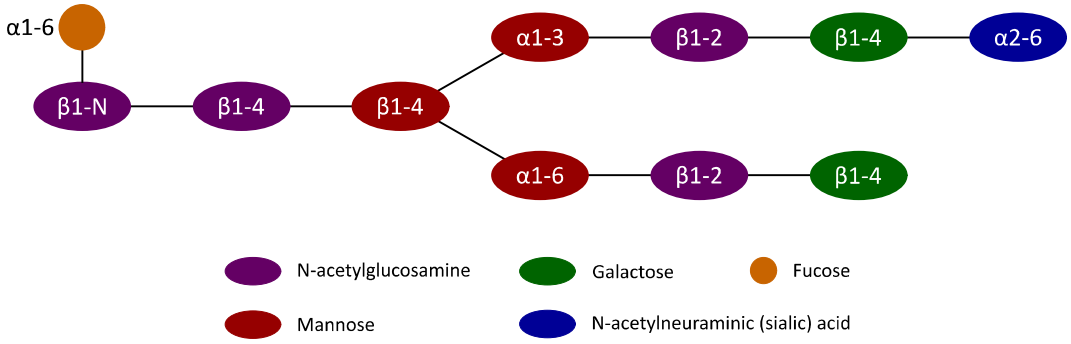


Figure 1.8: The major N-linked glycan structure of VWF The major N-linked glycan structure is comprised of N-acetylglucosamine, mannose, galactose, sialic acid and fucose, the colours of which are indicated.

In contrast to other plasma glycoproteins, N-linked glycans can also carry ABO(H) blood group determinants, with these being identified on approximately 15% of N-linked carbohydrates in plasma VWF. These determinants were not located to specific N-glycan sites, however, ABO(H) determinants were concentrated at the two largest structures, at N1515 and N1574[134]. The antigens of blood groups A, B and O (A, B and H, respectively) consist of complex carbohydrate molecules, with the addition of N-acetyl-D-galactosamine or D-galactose to the common H-determinant yielding type A or B antigens, respectively[138, 143, 144]. This carbohydrate addition or lack of, in turn, determines blood type O, A, B or AB. Interestingly, there is a strong correlation between blood group and plasma VWF levels, with those in blood group O displaying the lowest plasma VWF levels and blood group AB displaying the highest levels (AB>A>B>O)[145]. Blood group O display plasma VWF levels 25% lower than in non-O individuals. Whilst there is not a confirmed explanation for this correlation, the most likely hypothesis is blood type determinants having an influence on VWF clearance[146]. This would implicate that blood type antigens alter susceptibility to ADAMTS13 cleavage, with AB being the least susceptible and O being the most[147]. The fact that blood type determinants are concentrated at two asparagine residues within the A2 domain would support the theory of these influencing AMATS13 cleavage.

Following N-glycosylation, monomers undergo C-terminal tail-to-tail dimerisation between adjacent CK domains[148-150]. Crystal structures of the CK domain reveal intra-domain disulphide bonds between eight of the eleven cysteines within this domain, leaving three cysteines (C2771, C2773, C2811) to dimerise with an adjacent CK domain[43]. However, rather than

disulphide bonds forming between alike residues (C2771-C2771), EM findings suggest an antiparallel arrangement between the long axis of the two monomers, meaning C2771 of monomer 1 disulphide bonds with C2773 from monomer 2, and vice versa[151]. Dimerisation is initiated by PDI-1, which forms the first two disulphide bonds, with the final C2811-C2811 bond left to form to most likely protect the first two bonds, rendering dimerisation irreversible[151, 152]. Translocation of VWF from the ER is determined by dimerisation and glycosylation[153].

1.5.2. Glycan Modifications in the Golgi network

Once VWF enters the Golgi it undergoes further modification. The N-linked glycans undergo tyrosine sulphation at five sites (N1515, N2223, N2290, N2400 and N2790) which are preceded by Pro-XXX-Arg/Lys/His motifs[141, 154, 155]. Multimers also undergo O-linked glycosylation, the attachment of a sugar molecule to the oxygen atom of serine or threonine, followed by other carbohydrates such as galactose and sialic acid[156]. These O-linked glycan structures (Figure 1.9), along with N-glycans, make up 20% of the final monomeric mass of VWF[157]. Whilst N-linked glycans are dispersed throughout the monomer, O-linked glycans are far more clustered with eight flanking the A1 domain (T1248, T1255, T1256, S1263, T1468, T1477, S1486, and T1487) and the remaining two in the A3 (T1679) and C1 (T2298) domains[158]. The structure of these also varies greatly from N-linked glycans, being characterised by short mucin-type carbohydrates. Through mass spectrometry analysis, a disialylated core 1 structure, known as the Thomas Friedenreich antigen, was identified as the most common O-linked glycan structure, accounting for approximately 70% of O-linked glycans[156, 159].

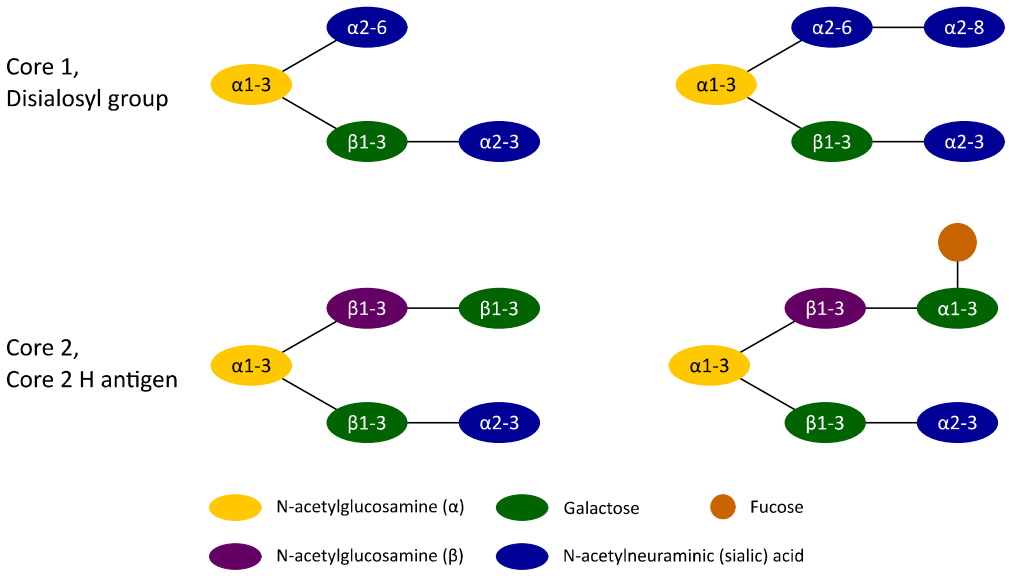


Figure 1.9: O-linked glycan structures Displayed are the O-linked glycan structures attached in the Golgi. These are comprised of N-acetylglucosamine, galactose, sialic acid and fucose, the colours of which are indicated.

Considering the location of eight O-linked glycosylation site surrounding the A1 domain (Figure 1.10), which houses the platelet GPIbα binding site, glycans at these sites influence platelet binding affinity, with mutations at these sites resulting in an increased or decreased binding affinity[160, 161]. Mutations in the hinge-linker region between the D3 and A1 domains show increased platelet capture under shear stress, likely due to increased flexibility and accessibility to the GPIbα binding site rather than glycans affecting the VWF A1-GPIbα bond[162]. Similarly to N-linked glycans, O-linked carbohydrates carry ABO(H) blood group determinants and also influence plasma VWF levels[156, 161, 163], however, are dispensable for multimerisation[161]. Therefore, whilst influencing processes such as platelet binding and VWF clearance, O-linked glycans are not wholly essential in the biology of VWF which agrees with the fact that no mutations resulting in VWD, a genetic disorder caused by missing or defective VWF, have been reported.

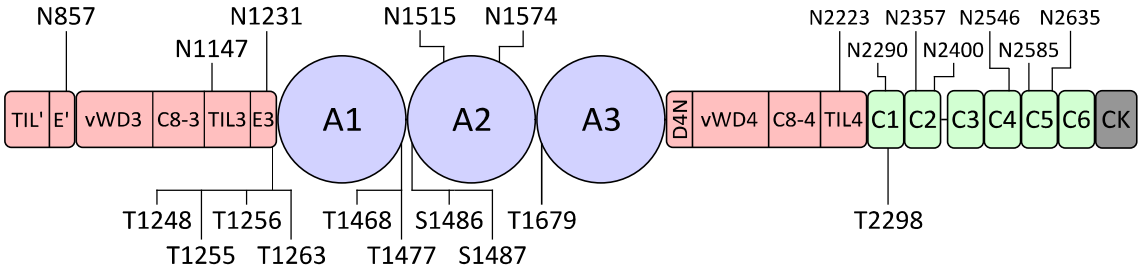


Figure 1.10: Mapping the N- and O-linked glycans in the VWF monomer Indicated are the N- and O-linked glycosylation sites in the VWF monomer. N-linked sites are indicated above at asparagine residues and O-linked sites are shown below at serine and threonine residues[140].

1.5.3. Multimerisation in the Golgi

Whilst there appears to be no limit to the size of the protein when transported from the Golgi to the cell surface, that is not the case for translocation between the ER and the Golgi with the relative cargo vessels having limited capacity. The majority of oligomeric proteins undergo polymerisation in the lumen of the ER as it is well suited for disulphide bond formation, being at a neutral pH (pH 7.4) and containing suitable chaperones, oxidoreductases and redox potential. The Golgi, on the other hand, lacks these and is relatively acidic (pH 6.2), conditions which generally prevent disulphide bond formation[164]. VWF head-to-head multimerisation has even been shown to be optimal at a more acidic pH than the Golgi, at pH 5.8, conditions which are typically found in WPBs[165]. However, neutralising the pH in the Golgi by addition of ammonium chloride, chloroquine or monensin abolishes multimer assembly[135], showing the importance of acidity in the specific mechanism that VWF has adopted[132].

The VWF propeptide is cleaved by furin at Arg-Ser-Lys-Arg at position 760-763, cleaving following R763[166]. This can occur before, during and following multimerisation. Histidine residues are regularly employed by proteins as pH sensors and with a pK_a of about 6 in the chemical environment of the Golgi, histidine is ideal to detect this change in pH[167]. Furin itself uses an autoinhibitory mechanism, as it is only able to cleave other substrates when His69 detects a change in pH, becoming protonated and causing its own propeptide to dissociate[168]. Within VWF, the D1D2-D'D3 segment was sufficient for pH-dependent multimerisation and tubular storage, therefore, the pH-sensing histidine residues must be located within this region[169]. VWF mutations to alanine at positions H395 and H460 in D2 caused a defect at inter-dimer disulphide bonds, preventing the acquisition of a positive charge as occurs with histidine. Replacement of the H460A mutation to other positively charged amino acids (arginine or lysine) restored the ability of D'D3 to multimerise, showing the importance of a positive charge at this position[170]. This is consistent with the VWF propeptide forming non-covalent homodimers with the D'D3 region, positioning two D'D3 regions in close proximity and subsequently facilitating disulphide bond formation[171]. It is suggested that H395 regulates dimerisation by its interaction with Y610 which ordinarily form a T-shaped aromatic interaction but converts to a stronger π-cation bond upon protonation[63].

The propeptide is employed as a disulphide oxidoreductase and remains non-covalently associated with VWF dimers and multimers, promoted by the high Ca^{2+} concentration and acidic pH of 6.2 experienced in the Golgi[165]. It remains associated in order to catalyse multimerisation using its vicinal cysteine CGLC motifs at positions 159-162 (in D1) and 521-524 (in D2), similar to that employed by disulphide isomerases. These motifs have been shown to be essential, with the insertion of glycine residue in the middle of either to form CGGLC inhibiting multimerisation[172]. Interestingly, studies have shown that the propeptide can be synthesised in parallel and still catalyse VWF multimerisation[148], however, upon deletion of either or both domains of the propeptide, solely dimers form[173]. Cleavage of the propeptide, on the other hand, has been shown to be dispensable for multimerisation as an R763G phenotype presents normal multimers[174], however, this mutation also appears to affect proteolytic processing of VWF and consequently FVIII binding[175].

During multimerisation, disulphide bonds form between adjacent D3 domains (Figure 1.11). The propeptide forms an intra-chain disulphide-linked intermediate with the D'D3 region in the ER, as is consistent with the oxidoreductase model. This intermediate rearranges in the Golgi and directly leads to the formation of VWF multimers by inter-dimeric disulphide bonds[169]. Flexibility between D3 subdomains is of particular importance concerning

multimerisation. The lack of disulphide bonds across the VWFD-C8 or C8-TIL boundaries allows structural rearrangement and exposure of D3 cysteines responsible for multimerisation[63]. Two cysteine residues in D3, C1099 and C1142, were identified as those which participate in multimerisation, with a mutation of either to alanine resulting in a loss of multimers[176, 177]. These cysteine residues are buried in hydrophobic pockets yet lie near the solvent accessible surface, protecting them from undergoing multimerisation in the ER. In the first helical turn of the C8-3 α-helix is C1099, buried by side chains including VWFD3 residues M872 and V919. Meanwhile, C1142 is buried by the bulky side chains of TIL3 residues L1135, Y1140 and W1144[63]. Unlike dimerisation in the CK domain, adjacent D3 domains lie parallel to each other with both inter-subunit disulphide bonds being formed between alike residues, i.e. C1099-C1099[177].

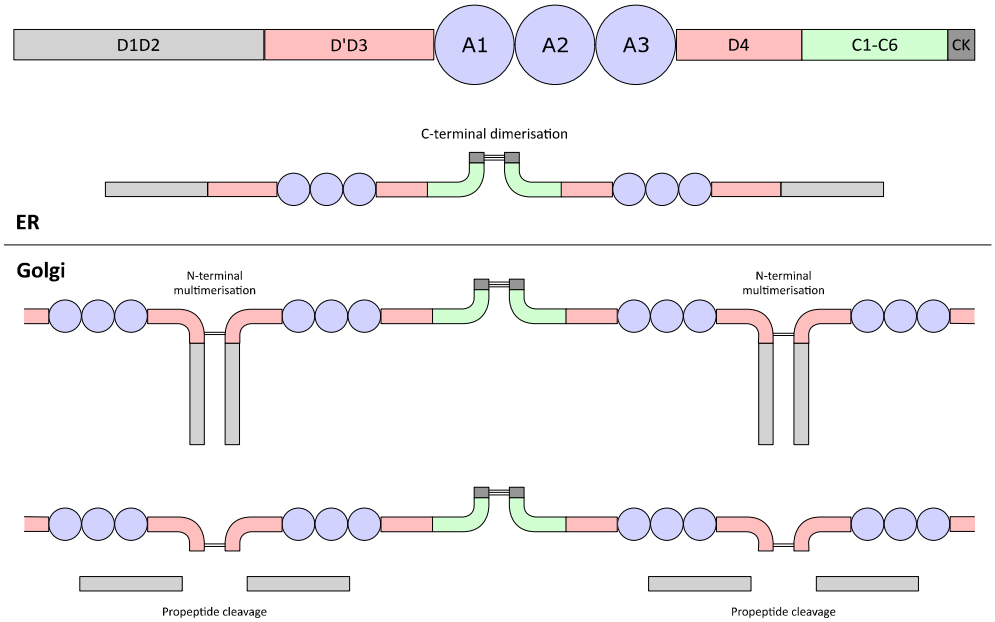


Figure 1.11: Dimerisation and multimerisation of VWF Displayed is the method of dimerisation of VWF, in the ER between CK domains, and multimerisation, in the Golgi between D3 domains. Colours corresponding to the different domains are indicated above. Disulphide bonds between monomers are indicated with solid black lines.

1.5.4. Storage and secretion

VWF produced in the endothelial cells is either stored in WPBs for stimulated release[178] or released constitutively into the blood plasma and subendothelial matrix[27, 179]**.** Constitutive secretion is a continuous release of low molecular weight multimers using anterograde carriers maintaining a baseline level of VWF in circulation and accounts for around 95% of secreted VWF[180]. When released into the blood plasma, it serves to bridge the gap between platelets and the extracellular matrix. VWF released into the subendothelium associates with Type VI collagen microfibrils[181].

The remaining VWF is stored in WPBs which are unique to vascular endothelial cells. They were first identified in 1964 by EM as a rod-shaped cytoplasmic component that was found in the endothelial cells of various organs[182, 183]. Whilst their function was initially unknown, they were soon shown to contain ultra-large (UL) VWF multimers[24, 182] along with a variety of other proteins residing throughout their structure, such as P-selectin, tissue plasminogen activator (tPA) and osteoprotegerin[184-187]. They have a diameter of 0.1-0.3 µm and a length of 1-5 µm, consisting of electron dense tubules with the propeptide, stored in a 1:1 ratio, and D'D3 domains forming the tubule and all domains C-terminal to D3 forming the surrounding matrix[27, 179, 188]. The wall has an outside diameter of ~25 nm and an inside diameter of ~12 nm, with tubules surrounded by a less dense matrix with parallel bundles, all housed in a lipid bilayer[171, 188].

1.5.4.1. VWF Storage in WPBs

As multimers enter the *trans*-Golgi, each dimer rearranges into a "dimeric bouquet" structure, triggered by a drop in pH from pH 7.4 (found in the ER) to 6.2 and an increased Ca^{2+} concentration. The C-terminal ~1350 residues of each subunit zip up with the CK to A2 domains acting as the "stalk" and the CK domains linked at one end at an angle of approximately 130° [37], and the amino-terminal D'D3-A1 forming the "flower"[189-191] (Figure 1.13). Both EM (Figure 1.12) and negative staining studies have displayed this structure in vitro[191]. This allows the presentation of an adjacent multimer and results in efficient multimerisation. As this process continues, forming UL multimers, and due to the luminal conditions of the trans-Golgi network, these bouquet structures arrange into a right-handed helical structure, with the "flower" D1D2-D'D3 region forming the core of the tubule[192, 193]. The large size of the D1, D2 and D3 assemblies, each around 380 residues, provides a framework for the formation of these helical tubules[63]. This method also allows for the long strings of covalently linked VWF to form without entanglement and allows for rapid unfurling when secreted, exposing multimers to a neutral pH[193].

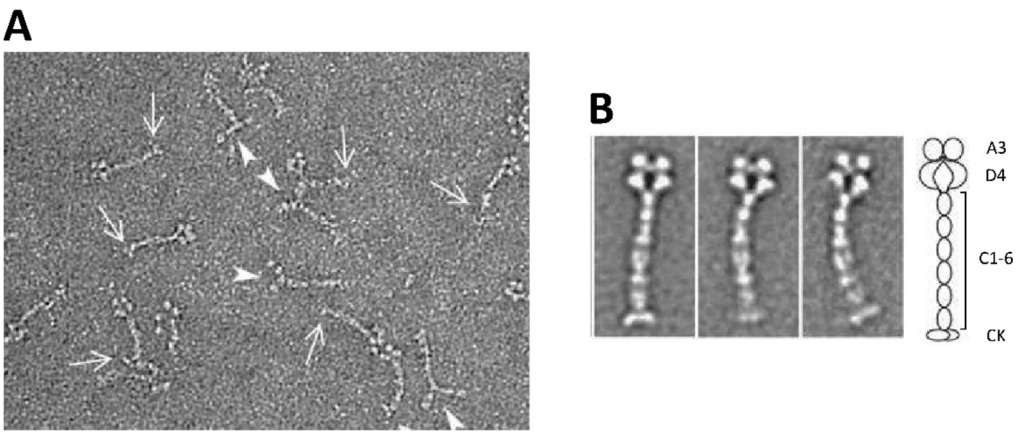


Figure 1.12: EM studies of VWF A3-CK bouquet structure Displayed in images A and B are the bouquet structures of a truncated form of VWF (A3-CK) in EM studies. In image A, arrows mark the CK domains which bind in dimerisation, with arrowheads indicating globular head domains which have separated[191].

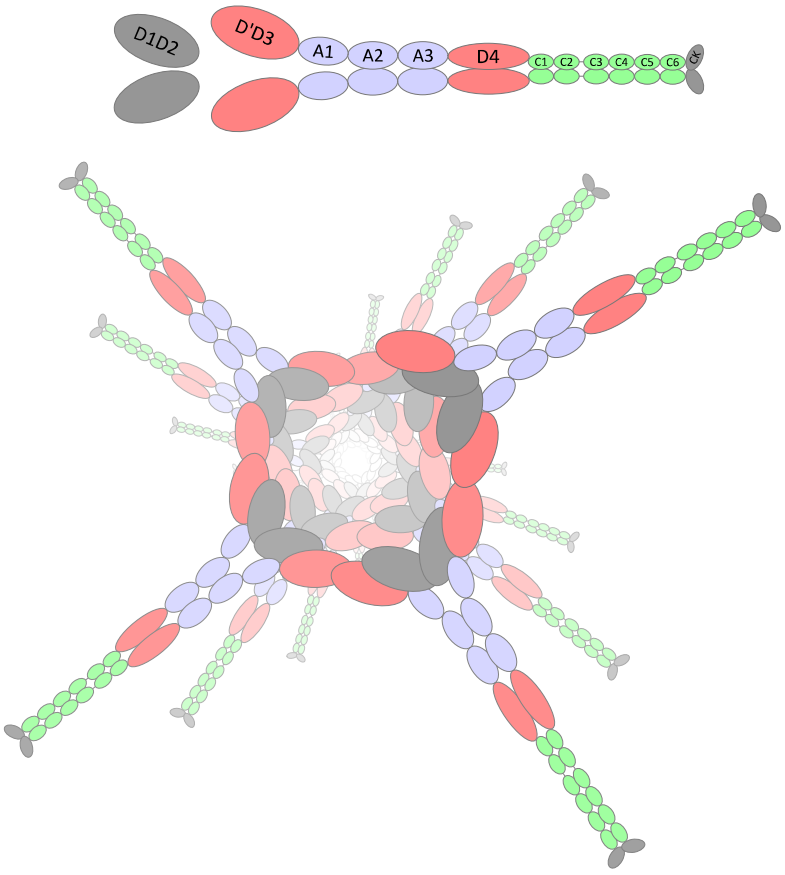


Figure 1.13: VWF bouquet structure and multimerisation Displayed above is the "dimeric bouquet" formed by VWF caused by the luminal milieu of the trans-Golgi. The A2 to CK domains form the "stalk" and the D'D3 region forms the "flower". Multimers arrange into a right-handed helical structure in the trans-Golgi with 4.2 units (dimers) per turn with the D'D3 region and D1D2 propeptide lining the lumen of the helix.

Storage of VWF multimers in WPBs represents an interesting phenomenon, with both requiring the presence of the other to be produced, as VWF-deficient animals do not contain

WPBs[148]. Multimers formed in the Golgi arrange into tubules and are incorporated into vesicles by a clathrin/AP-1 coat, which protrude from the trans-Golgi network and act to mediate transport between subcellular compartments by forming small vesicular carriers[188, 194]. Where coat function is experimentally prevented, WPBs cannot form, leading to small VWF-containing puncta being observed and regulated secretion being significantly depleted. Two AP-1 effectors, aftiphilin and γ-synergin, have also been shown to be required for the regulated secretory phenotype, as knockdown of these two proteins leads to VWF being released in an unregulated manner, resulting in a 3.4-fold increase in active VWF multimers in circulation. With VWF maturation occurring within WPBs, this may lead to insufficient time for UL multimers to be formed and a plasma VWF pattern dominated by low molecular weight multimers[194].

Once immature WPBs bud off from the trans-Golgi network, they undergo a maturation process whereby they form their typical cylindrical shape[195]. This cylindrical shape relies upon the formation of multimers as mutations that affect VWF multimerisation, such as Y87S in the D1 domain, result in the formation of spherical storage granules[193]. As WPBs disperse throughout the cytoplasm, the clathrin/AP-1 coat, aftiphilin and γ-synergin are shed[194, 196] and multimers undergo further multimerisation to become UL, reaching up to 20,000 kDa[190]. Maturation also leads to a further drop in pH to around pH 5.4[197], causing multimers to be tightly packed into their helical structure[27, 179], with the D'D3 domains and propeptide forming the core of the tubule and the remainder of each monomer protruding outwards, with each dimer spanning ~120 nm in length when extended[198]. Maturation coincides with the acquisition of a number of membrane proteins, including Slp4-a, MyRIP and several small GTPases, including Rab27A, Ra3B and Rab3D, all of which have been shown to have an essential role in the formation of WPBs and controlled VWF secretion[188, 199, 200]. Rab27A and Rab3B have been shown to be involved in the recruitment of the Rab27A effector Slp4-a[201]. MyRIP and Slp4-a appear to have opposing effects on WPB exocytosis with Slp4-a acting as a positive regulator while MyRIP acts as a negative regulator[202]. They both bind to Rab27A, however, Slp4-a has a competitive advantage over MyRIP as Slp4-a can bind to both the GTP and GDP bound form of Rab27A, whereas MyRIP can only bind to GTP-Rab27A[203, 204]. This gives the probability of release favouring exocytosis[201]. The organisation of VWF tubules changes between immature and mature WPBs, with tubules being curved and disorganised with large separation in the former, but regularly arranged, running longitudinally and far more electron dense in the latter[188, 195, 196, 205]. Interestingly, whilst multimerisation is essential for normal VWF function, it is not required for the formation of tubules or WPBs[171, 193].

UL multimers can be tightly packed much more efficiently when in helices with a ~45-fold compaction of VWF compared to its globular form, as it circulates in plasma[198]. Helices consist of 4.2 repeating units per 11 nm turn (one dimer per repeating unit), suggesting that every 1 µm of tubule length corresponds to ~780 VWF subunits, a mass of ~195 million Daltons and, with each monomer being ~60 nm in length, an extended length of ~47 µm[171]. However, as multimer gel analysis has shown that a VWF multimer consists of up to 50 dimers equating to a 130 nm tubule and WPBs can be up to 5 µm in length, they must house multiple tubules packed in a paracrystalline order by a pH-sensitive mechanism[132, 189, 195]. Electron cryotomography has shown that these tubules are flexible, bending at the end of the granules to accommodate the membrane curvature[195]. The tips of tubules regularly touch the membrane of the WPB, however, tubules appear to remain a constant length away from the side membrane along the entire length[188, 196].

The WPB contents are primarily involved in haemostasis and inflammation, responding its complex surrounding environment. Some WPB co-residents directly interact with VWF which likely means they are trafficked into the storage vesicles. FVIII, the clotting factor which circulates bound to VWF in circulation, is also stored alongside VWF in WPBs[206]. It is produced in endothelial cells alongside VWF and resides in the lumen of the helical tubules, bound to its site in the D' domain[62]. The trafficking of FVIII from the Golgi to WPBs has been shown to be VWF-dependent. The multiple co-residents indicate the extensive role that WPBs play in not only haemostasis, but also inflammation, haemodynamics and angiogenesis, allowing the endothelium to an array of changes in its microenvironment[187]. P-selectin is a cell surface glycoprotein which regulates inflammation acting as a cell adhesion molecule on the surface of endothelial cells or platelets and interacts with the D'D3 domains of VWF, residing in the lumen of the UL VWF helical multimer[207, 208]. A number of inflammatory regulating co-residents are stored in WPBs including osteoprotegerin, interleukin-8, angiopoietin-2 (Ang-2), CD63, α1,3-fucosyltransferase VI and eotaxin-3[209-217]. Other co-residents induce a variety of responses, such as vasoconstriction by endothelin-1 and endothelin-converting enzyme[218], direct fibrinolysis by tPA[219] and apoptosis by galectins-1 and -3[185]. Some of these interact directly with VWF, such as osteoprotegerin with the A1 domain and galectins-1 and -3 with N-linked glycans[220], meaning they are likely trafficked into WPBs[185, 210]. These proteins are incorporated by different mechanisms and at different stages of WPB development, for example, P-selectin is recruited to forming WPB in the trans-Golgi whereas CD63 is recruited to already budded WPBs in the post-Golgi by an AP-3-dependent route[215]. Other proteins, such as interleukin-8 and tPA, are incorporated as a result of a random inclusion process meaning all

WPBs are different and unique[154, 212, 221]. Whilst multiple studies have shown the wide-ranging contents of various WPB co-residents, some studies have shown that this is by mechanisms more than a random inclusion process but by dynamic regulation of WPB contents. Studies have shown that, while interleukin-8 is included by a random inclusion process, its synthesis is in response to inflammatory mediators such as interleukin-1β[212] and eotaxin-3 is only produced following endothelial cell stimulation by interleukin-4[217].

1.5.4.2. VWF Secretion

As mentioned previously, the majority of VWF is released as low molecular weight multimers through constitutive secretion without stimulation[27, 179, 180]. Although the greater proportion of circulating VWF is released by constitutive secretion, secretion from WPBs is more significant in terms of haemostasis, with regulated secretion releasing more haemostatically-active, high molecular weight multimers. UL VWF multimers are released from WPBs by regulated secretion (Figure 1.14). Regulated secretion occurs either by injury[222] or stimulation by secretagogues, leading to a rise of cytosolic Ca^{2+} or cAMP[223]. Secretagogues include thrombin, histamine (both Ca^{2+}-raising), adrenaline, serotonin and vasopressin (cAMP-raising)[224]. A rise in cytosolic Ca^{2+} tends to be by agonists that are mediators of inflammation or thrombosis in a local manner and involves both peripheral and central granules[225-227]. These Ca^{2+}-raising agents are most likely mediated by calmodulin, with thrombin and histamine particularly resulting in the activation of the small GTPase RhoA[228-230]. This causes stress fibre formation and cell retraction, resulting in exposure of the subendothelium by disassembly of tight and adherens junctions which limit paracellular permeability and mechanically couple neighbouring cells, respectively[227, 231]. The organisation and dynamics of the actin cytoskeleton is effectively modulated by RhoA and these disruptions leads to severe subcutaneous and whole-body cavity oedema formation[227, 230, 232]. Other Ca^{2+}-raising secretagogues include peptido-leukotrienes, superoxide anion, VEGF and ceramide, with purine nucleotides causing an influx of both Ca^{2+} and cAMP[187].

Secretagogues which induce cAMP influx, however, solely involve peripheral granules and are more likely to be responsible for regulating plasma VWF levels[225-227]. They cause disruption of stress fibres, strengthening of the cortical actin rim and tightening of cell-cell contacts[227]. Cytosolic cAMP activates Rap1 which induces multiple signalling cascades, most notably inhibition of RhoA which decreases radial stress fibres and activates Cdc42 to increase junctional actin[41]. Rap1 activation also controls vascular endothelium (VE)-cadherin-mediated cell junction formation and induces reorganisation of the actin cytoskeleton[233]. Vasopressin analog desmopressin (DDAVP) is used in the treatment of patients with VWD and

Haemophilia A, increasing plasma levels of VWF, FVIII and tPA. It binds to the vasopressin V2 receptor on endothelial cells, leading to the activation of trimeric G-protein and adenylyl cyclase, causing production of cAMP, the activation of protein kinase A and exocytosis of the WPB[225]. Both methods of initiation lead to fusion of the WPB wall to the endothelial cell membrane and the formation of a secretory pore[234]. The WPB begins to partially hydrate and become more alkaline, causing the core to swell due to a loss of alignment in the bouquet structure[154, 195, 197]. The tubular nature of VWF storage in WPBs enables rapid unfurling of UL multimers upon secretion, with this occurring within 15s following stimulation in mice[193, 235].

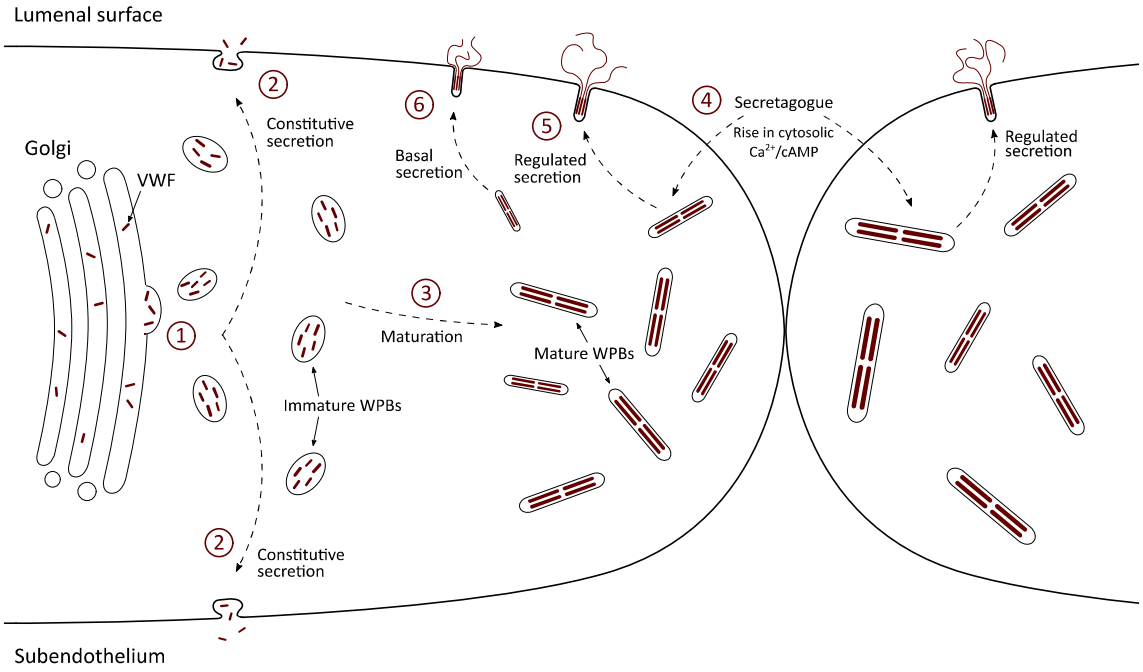


Figure 1.14: WPB formation and VWF secretion pathways WPBs are formed by VWF multimers budding off following processing in the Golgi apparatus (1). Multimers can be released by constitutive secretion into both the blood vessel lumen and subendothelium (2), as occurs for ~95% of circulating VWF multimers. Alternatively, WPBs undergo maturation with multimers joining together to form UL multimers (3). Stimulated release of UL multimers occurs when secretagogues cause a rise in cytosolic Ca^{2+} or cAMP (4), stimulating the fusion of WPBs with the cell membrane. This causes release of the WPB contents into the vessel lumen (5). The final method of VWF release is basal secretion (6) where WPBs spontaneously fuse with the endothelium[236].

UL VWF multimers released by WPBs rarely circulate in the blood as they are rapidly cleaved by the disintegrin and metalloproteinase ADAMTS13. When UL multimers are released, they bind to the exterior membrane of endothelial cells via P-selectin in the D'D3 region which in turn binds to the surface of the endothelial cell[207]. As these multimers are exposed to shear stress from rapid blood flow they elongate, some being several millimetres long, exposing the ADAMTS13 cleavage site (Y1605-M1606) in the A2 domain and causing cleavage into smaller multimers[182, 237]. The cleaved multimers, however, are still UL in size though UL multimers are

rarely found in circulation, suggesting they undergo rapid ADAMTS13-mediated proteolysis in circulation[238]. P-selectin, meanwhile, will remain on the surface of the activated endothelial cell, acting as a cell adhesion molecule in inflammation[239]. UL multimers are only seen transiently in plasma VWF, following treatment with DDAVP or in diseases such as thrombocytopenic purpura[20, 182, 240]. Once released into the neutral pH 7.4 of the blood vessel or subendothelium, VWF circulates in its disorganised globular form as a "folded spring", allowing it to be highly dynamic when required for haemostasis[241-243]. In around 25% of instances, the sudden exposure of the WPB contents to an alkaline pH of the blood by the insertion of a fusion pore about 12 nm in diameter can cause the collapse of the helical VWF structure. It is then retained in the cell alongside its propeptide and P-selectin, while smaller molecules such as interleukin-8 and CD63 are released. This is known as a "lingering kiss" and demonstrates that WPB exocytosis is not inextricably associated with VWF release and can be selective[244]. Upon secretion and exposure to a neutral pH, the unbound cysteines at either end of the concatemer, C1099 and C1142 which are otherwise involved in multimerisation, become buried within the structure as occurs in the ER in order to prevent spontaneous multimerisation[63]. There is also evidence of a third secretory method, a combination of both constitutive and regulated secretion with proteins targeted to storage granules but released without provocation, known as basal secretion[236]. WPBs can move freely about the cytoplasm with some inevitably situated at the cell periphery, allowing fusion of the WPB with the cell wall and release of its contents into either the lumen or the subendothelium[154, 236, 245, 246].

1.5.5. Platelet-derived VWF

Platelet-derived VWF, produced in megakaryocytes and transported into platelets, comprises 20% of plasma VWF and is rich in high molecular weight multimers[247]. Platelet-derived VWF, in contrast to endothelial-derived, is solely stored in α-granules and not secreted constitutively[248]. Whilst endothelial-derived VWF is the dominant form of plasma VWF as it binds more efficiently to platelet GPIbα[27, 179, 249], that secreted from platelet α-granules has been shown to have a role in haemostasis, partially correcting the abnormalities in pigs severely lacking endothelial-derived VWF[23]. In contrast to WPBs where VWF is the primary storage molecule, α-granules house an array of other components such as fibrinogen and coagulation factors V, XI and XIII[248]. VWF is stored in a peripheral electron-lucent zone[250] alongside co-inhabitants also found in WPBs, such as P-selectin[248, 251]. The contents of α-granules in mature platelets varies greatly with some staining for a lack of fibrinogen and others for a lack of VWF, showing that packaging of granules is not a regular process and all granules are unique[252]. These α-granules also differ in shape to WPBs being more spherical,

however, VWF multimers are organised into clusters of tubules, similar to that found in WPBs[171].

The synthesis, storage and secretion of platelet-derived VWF is largely similar to that of endothelial cells, from dimerisation in the ER, multimerisation in the Golgi and translocation into storage vesicles. Coat proteins (clathrin), adaptor proteins (AP-1, AP-2 and AP-3) and soluble GTPases are all essential for the trafficking and maturation of α-granules[248]. Upon megakaryopoiesis, α-granules are distributed to platelets with organelles delivered to proplatelets along microtubule bundles[253]. Platelet activation occurs following an endothelial lesion, where platelets come into contact with exposed collagen and VWF. This causes exocytosis of platelet storage granules, releasing their contents, importantly, concentrated to the site of vessel lesion[254]. Endothelial-derived VWF, however, has been shown to be the dominant form of VWF in primary haemostasis. Although a lack of platelet α-granules in congenital gray platelet syndrome has been shown to lead to a mild-to-moderate bleeding disorder, this is due to more than an absence of platelet-derived VWF. Other factors include a lack of pro-adhesive proteins which are ordinarily stored in α-granules[255].

1.5.6. Multimer distribution and VWF self-association

Circulating VWF exists in a variety of molecular weight forms, high, intermediate and low molecular weight multimers. These groups are separated by high molecular weight multimers between 5500 and 10,000 kDa, intermediate 3000-5000 kDa and low 500-2500 kDa, with UL multimers stored in WPBs over 10,000 kDa. In terms of primary haemostasis, high molecular weight multimers are the most effective with their abundance of collagen, GPIbα and platelet aggregation-inducing binding sites[149, 256]. The size of circulating VWF multimers are regulated by both ADAMTS13 cleavage and trimeric glycoprotein thrombospondin-1 (TSP-1) which acts as a disulphide bond reductase/isomerase and the cleavage. TSP-1 is stored in a number of cell types including endothelial cells and smooth muscle cells[257, 258], however, is stored in abundance alongside VWF in platelet α-granules and is released upon platelet activation with its presence in circulating plasma at very low concentrations[259].

The active site for reductase/isomerase activity on TSP-1 has been shown to be a CGXC motif with these cysteines alternating between free thiols and disulphide bonded when coordinated with a protein substrate[260]. This cycle causes the formation, reduction or rearrangement of disulphide bonds in that protein substrate. In this case, VWF is that protein substrate with disulphide bond reduction, formation and rearrangement occurring in circulation. The interaction between VWF and TSP-1 begins between the TSP-1 type I domain and the VWF A3

domain which subsequently allows the interaction of the TSP-1 C-terminus with VWF disulphide bonds. This reduction of disulphide bonds has been shown to favour high molecular weight multimers, likely due to their higher avidity compared to small multimers[261]. This reduction and rearrangement of disulphide bonds explains the presence of free thiols in plasma VWF with their physiological significance becoming ever more apparent.

TSP-1 appears to influence plasma and platelet-derived VWF in different ways having been shown to compete for binding with ADAMTS13 in the VWF A2 domain of platelet-derived VWF[262]. This is significant when considering UL multimers released in response to platelet activation, however, as endothelial-derived VWF is sufficient to maintain normal haemostasis, TSP-1-deficient mice do not exhibit a haemostatic disorder[263]. Whilst it was also suggested that TSP-1 can mediate platelet binding in the absence of VWF, research has shown this not to be the case although TSP-1 does require VWF to modulate the rate of thrombus growth in injured arterioles[264]. The interaction of TSP-1 with the CD36 receptor on the surface of endothelial cells provides a feed-forward loop that enforces the interaction of platelets in a thrombus formation[265]. This is demonstrated by TSP-1 knockout mice exhibiting an impairment in thrombus formation[264].

1.5.6.1. VWF self-association

VWF self-association is a phenomenon which has emerged over the last 20 years, first reported where thrombus formation was promoted when wild-type (WT) VWF perfused over immobilised VWF lacking the functional GpIbα binding site, resulting in VWF aggregation[266]. The importance of free thiols has been demonstrated where blockade of free thiol groups significantly reduced VWF-mediated platelet recruitment under physiological flow conditions[267, 268]. This process has been shown to occur under fluid shear stress where VWF undergoes a conformational change from a loosely coiled ball to an elongated structure[269]. Multimers form into a flexible clump when perfused over a collagen surface, forming a network of interconnected fibres[270]. Self-association promotes binding of several VWF multimers onto a single platelet GpIbα receptor, triggering mechanotransduction and platelet activation[271]. UL VWF multimers undergo self-association with each other upon stimulated release, stretching millimetres into the bloodstream whilst also recruiting circulating globular VWF[182]. Two mechanisms of VWF self-association have been proposed, by reversible ionic interactions and stable, covalent interactions[266, 269]. Those by covalent interactions occurred between circulating multimers and secreted UL multimers. Unlike circulating VWF multimers, free thiols found in UL multimers are “buried” but are exposed when secreted and stretched by fluid shear stress[272].

Studies have implicated a number of mechanisms that impact and regulate the role of VWF self-association. At least part of self-association has been shown to not rely upon disulphide bridging[269], however, the role of unpaired cysteines has been heavily implicated. Formation of VWF strings is inhibited by thiol alkylation, highlighting the utilisation of thiol/disulphide exchange[268]. Formation of disulphide bonds is functionally critical in binding to coagulation FVIII, GpIbα and folding of the VWF A2 domain[273], however, recent research has shown that not all cysteines are involved in disulphide bonds with both recombinant and plasma VWF containing free thiols[274]. Those cysteines found as free thiols, however, more often than not exist in their disulphide-bonded state[267]. This shows that the formation of free thiols is a dynamic process following the reduction of certain disulphide bonds by oxidoreductases, in particular TSP-1[275]. Proteolytic enzyme ADAMTS13 has also been shown to possess reducing capacity close to its proteolytic active site[276].

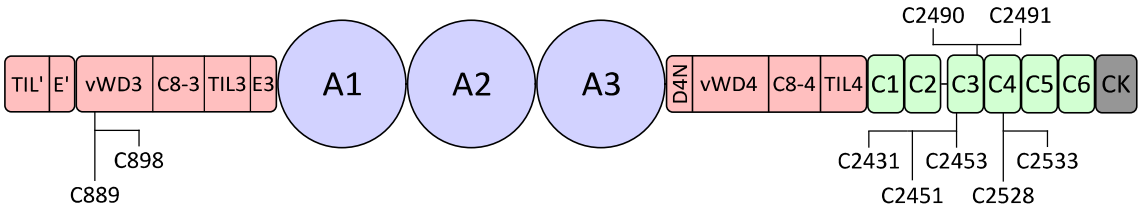


Figure 1.15: Mapping of the VWF free thiols Indicated are the locations of cysteine residues that have been identified as existing as free thiols in plasma VWF[267].

Free thiols are primarily localised to the D3 domain and C-terminal C domains. They are distributed amongst multimer sizes, however, larger multimers generally exhibit fewer free thiols. This can likely be explained by the reliance of low molecular weight multimers on aggregation to efficiently participate in platelet recruitment. Studies have shown a number of cysteines are found as free thiols in circulation (Figure 1.15). A particular cysteine-rich sequence, CGRC, found in both the C3 and C4 domains, contained one cysteine in a reduced state, suggesting isomerase activity at this sequence[267]. Cysteines thought to be of particular importance in promoting lateral self-association are C2451 and C2453 in the C3 domain. The C2431-C2453 disulphide is the most common disulphide bond to be reduced with C2453 the most common free thiol found in ~75% of cases, further implying that reduction of disulphide bonds occurs during secretion or circulation. Both C2431 and C2451 residues have been shown to form inter-chain disulphide bonds, forming trimers and higher order oligomers, with the reduction of these disulphides being the rate-limiting step[274]. Other cysteines found as free thiols are C889 and C898 in the D3 domain, C2448, C2490 and C2491 in the C3 domain and

C2528 and C2533 in the C4 domain[268]. In terms of the overall importance of free thiols, blockade of free thiols reduced platelet capture in a shear-dependent manner with platelet capture significantly reduced at 1500 s^{-1}, shear rates widely found in small arteries[277, 278]. Collagen binding, however, was unaffected upon free thiol blockade, suggesting free thiols enhance flexibility of VWF multimers and enable multimers to withstand high shear rates following collagen binding[277]. Whilst individual cysteines that exist as free thiols have been identified, the explanation as to why those cysteines in particular are susceptible to disulphide reduction and the effect of individual free thiols on the physiological function of VWF remains to be determined.

A major regulator of VWF self-association is the shear-dependent unfolding of the VWF A2 domain. This has been shown with, firstly, mutations that prevent domain unfolding reducing self-association and platelet activation, and secondly, those that destabilise the A2 domain and increase unfolding resulting in enhanced self-association. The A2 domain has also been shown to interact with the neighbouring A1 domain, regulating its activity. This displays a delicate balance within a single domain to regulate circulating VWF multimer size, through self-association and ADAMTS13 cleavage[279].

1.5.7. VWF clearance

Clearance of VWF multimers occurs through cleavage by ADAMTS13 at the Y1605-M1606 in the VWF A2 domain, named the scissile bond[280]. ADAMTS13 is part of the ADAMTS family of Zn^{2+}-dependent metalloproteases and is produced in the hepatic stellate cells and vascular endothelial cells[281, 282]. This protease is essential for modulating VWF activity with ADAMTS13 deficiency being one of the main dispositions in a patient developing thrombotic thrombocytopenic purpura, a condition where blood clots form in the small vessels of the vascular system[283]. Reduced levels of ADAMTS13 has also been correlated with increased risk of coronary heart disease and ischemic stroke[284, 285].

In circulation whilst in its globular form, VWF remains resistant to proteolysis until it elongates and exposes the scissile bond. This limiting factor is the primary mechanism in regulating ADAMTS13 activity with blocking of ADAMTS13 or proteolysis by serine proteases only observed in pathophysiological conditions[286, 287]. Control of ADAMTS13 function is also limited by the N-linked glycan modification at N1574 with a lack of a glycan structure at this residue increasing VWF A2 domain unfolding and susceptibility to cleavage[141]. A conformational change in VWF multimers occurs when exposed to mechanical shear stress in the vascular flow and/or upon binding of multimers to subendothelial collagen. This increased tensile force to

which UL multimers are more susceptible[288] exposes the ADAMTS13 cleavage site and results in multimer cleavage[281]. The half-life of VWF in circulation has been shown to vary between 9 and 15 hours[289]. Proteolysis by ADAMTS13 has been shown to be both Zn^{2+}- and Ca^{2+}-dependent, with Ca^{2+} providing structural integrity to an essential protease loop region[290]. Whilst proteolysis of VWF multimers only occurs when multimers are elongated, a proportion of ADAMTS13 (~3%) circulates in complex with globular VWF, most likely high molecular weight multimers[291]. Whilst still allowing a thrombus to form, this complex likely mediates the progression of thrombus formation as mice exhibiting a truncated form of ADAMTS13 continue to display normal multimer patterns, however, are more thrombogenic under shear conditions[292].

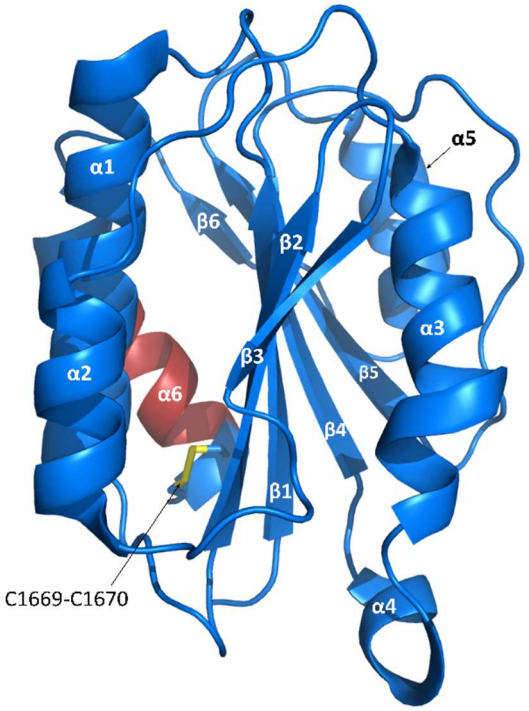


Figure 1.16: Structure of the VWF A2 domain Displayed is the structure of the VWF A2 domain with a central β-sheet region surrounded by six amphipathic α-helices. Each β-sheet and α-helix number is indicated. Highlighted in red is the A2 domain exosite (E1660-R1668) where ADAMTS13 binds and in yellow is the C1669-C1670 disulphide bond which prevents spontaneous domain unfolding[293].

The VWF A2 domain, where ADAMTS13 cleavage occurs, adopts a similar structure to the A1 and A3 domains with a central β-sheet surrounded by amphipathic α-helices (Figure 1.16). It differs, however, to its neighbouring domains as it does not possess an N-to-C termini disulphide bond or one of the α-helices, replaced by a less-ordered loop[293]. This allows the domain to be more susceptible to unfolding[281]. Unfolding occurs when shear forces are sufficient to cause water molecules to disturb the hydrophobic core leading to the loss of its tertiary structure. A unique C1669-C1670 disulphide bond has, however, been shown to be

vital in preventing spontaneous domain unfolding[294]. The ADAMTS13 spacer domain has been shown to bind to a key region in the VWF A2 domain, E1660-R1668, known as the VWF A2 domain exosite[295]. Whilst not exposed in the folded structure, the extraction of the molecular plug formed by the C1669-C1670 disulphide and the C-terminal α-helix exposes this site and allows ADAMTS13 binding[281]. Removal of the A2 domain exosite results in a dramatic 20-fold reduction in VWF cleavage[296]. A number of cysteine mutations have also resulted in increased VWF clearance in vivo, including mutations of C1130 and C1149 in the D3 domain and C2671 in the C6 domain. Whilst these mutations likely affect the structural integrity of VWF multimers, the significance of these cysteine residues in particular remains unclear[273].

Positioning of an A2 domain within a VWF multimer also affects its likelihood of undergoing cleavage. Tensile forces experienced within the multimer vary depending upon its distance from either end of the multimer[288]. Therefore, A2 domains in the centre of a multimer are on average more susceptible to cleavage than those towards the N- or C-terminus. Furthermore, an A2 domain in the centre of a high molecular weight multimer will be more susceptible to that in a low molecular weight multimer. This ensures that cleavage by ADAMTS13 favours high molecular weight multimers and, in the case of newly secreted VWF, UL multimers[281]. Higher shear rates lead to increased cleavage by ADAMTS13, as would be expected with a greater number of exposed A2 domain scissile bonds. Platelet-bound and secreted UL VWF, displaying a high shear rate due to their anchoring to subendothelial collagen and endothelial cells, respectively, have been shown to be particularly susceptible to ADAMTS13 cleavage[182, 297]. Reports have also suggested that CD36, the endothelial cell surface receptor that binds to TSP-1, is also able to bind to the ADAMTS13 TSP-1 repeat and localise ADAMTS13 to regulate cleavage of secreted UL VWF[298]. Whilst high shear rates result in A2 domain unfolding, a drop in shear rate has been shown to result in domain refolding, an important feature to prevent excessive cleavage[281, 288].

1.6. Secondary functions of VWF

Whilst VWF efficiently carries out its primary functions of anchoring platelets to the subendothelium and binding to FVIII in circulation, its influences in vascular biology are not limited to this. This includes roles in inflammation, angiogenesis and tumour metastasis with research increasingly showing VWF playing a vital role in these processes[239, 299, 300].

1.6.1. Inflammation

The role of VWF in inflammation was initially considered to be as an acute inflammatory biomarker. Being positioned at the interface between blood flow and tissues, endothelial cells play a key role in recruitment of leukocytes to areas of inflammation by altering their phenotype in response to inflammatory stimuli[301, 302]. This change results in the release of its WPB contents, the main storage molecule being VWF, allowing VWF to be used as an inflammatory marker[239, 302]. This method of diagnosis, however, has presented inconsistencies with elevated VWF levels not necessarily specific for an inflammatory response. With both mature VWF and VWF propeptide being secreted in equal molar concentrations but both displaying different half-lives, by analysing the ratio of mature VWF to the propeptide may allow differentiation between chronic and acute stimulation. Propeptide levels are only marginally increased by chronic stimulation whereas in response to an acute inflammatory response both mature VWF and the propeptide display similar increases[239, 303-305].

Understanding began to develop of an indirect impact of VWF on inflammation via P-selectin and other WPB co-residents (Figure 1.17). P-selectin is a cell adhesion molecule which recruits leukocytes to activated endothelium during inflammation, the first step in the cell adhesion cascade[306, 307]. The primary ligand that P-selectin binds to is P-selectin glycoprotein ligand 1 (PSGL-1), shown to be expressed on the surface of almost all leukocytes[308]. This allows chemokines located on the endothelial cell surface, such as monocyte chemoattractant protein-1 (MCP-1), to activate monocytes which in turn bind to other molecules involved in the cell adhesion cascade, such as intracellular adhesion molecule 1 (ICAM-1) or vascular cell adhesion molecule 1 (VCAM-1). Leukocytes thereby transmigrate into the arterial intima[309]. Whilst involved in inflammation, this mechanism also implicates P-selectin into atherosclerotic development. The binding site for P-selectin is in the D'D3 domain of VWF, being stored in WPBs and α-granules alongside VWF[207]. Upon activation of endothelial cells or platelets, secretion of VWF from WPBs along with its co-inhabitants leaves P-selectin bound to the cell surface. Soluble P-selectin is also found circulating in plasma and it has been shown that this is synthesised employing special mRNA which lacks the encoding sequence for the

transmembrane domain, rather than from shedding of membrane bound P-selectin[310]. Whilst the majority of circulating VWF is endothelial-derived (~80%), the majority of P-selectin is platelet-derived[311, 312].

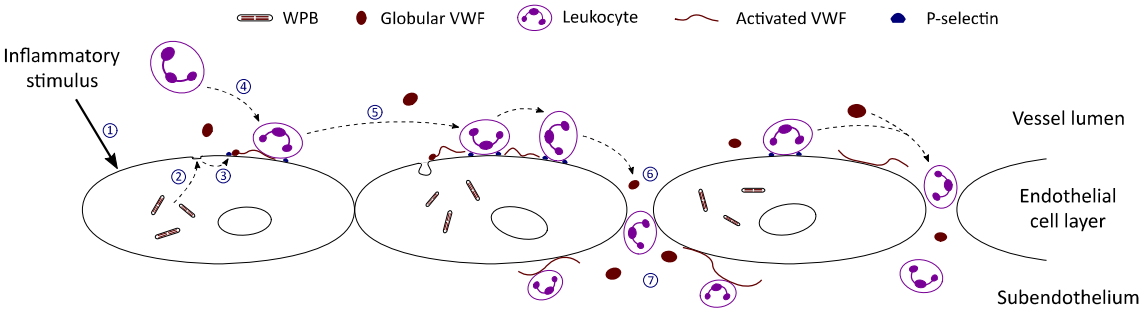


Figure 1.17: The role of VWF in inflammation An inflammatory stimulus on an endothelial cell (1) causes, amongst other effects, the secretion of WPB contents (2). This includes both VWF, a number of pro- and anti-inflammatory mediators and P-selectin, which remains on the endothelial surface (3). P-selectin recruits leukocytes to the active endothelium (4) and leads to leukocyte rolling (5). Endothelial junctions become weakened, leading to leukocytes and VWF migrating (6) and concentrating in the subendothelium (7).

The mechanism of co-inhabitation means VWF production has an indirect impact on inflammation, with it being shown that VWF-deficient mice present reduced leukocyte rolling as occurs in P-selectin deficient mice[313]. It was concluded that lack of WPBs due to the absence of VWF reduced expression of P-selectin on the endothelial surface, particularly in response to acute inflammatory stimuli. This was not the case, however, during sustained inflammation such as atherosclerosis, where P-selectin was potentially expressed on the endothelial surface from an adapted direct mechanism[314]. Any changes in inflammation may also be affected by altered expression of other inflammatory stimuli which are stored in WPBs such as interleukin-8, eotaxin-3 and Ang-2[212, 213, 217], and anti-inflammatory stimuli, such as osteoprotegerin which has been shown to block the RANKL/RANK pre-inflammatory mechanism in rheumatoid arthritis patients[209]. However, despite WPBs being heavily reliant on VWF production, their remarkable plasticity means they are not solely VWF-dependent, with selective cargo secretion being shown, suggesting these organelles have multiple roles[244]. This displays the essential role that WPBs play through their cargo and exocytosis in the development and progression of vascular disease.

Recent studies suggest that not only does VWF have an indirect impact on inflammation but also a direct impact on leukocyte transmigration into the peritoneum by weakening endothelial junctions[239, 315]. The mechanism by which VWF modulates permeability remains to be determined, whether by regulating the expression of the endothelial tight-junction protein claudin-5, or interaction with specific receptors such a $\alpha_v\beta3$ or LRP1[316, 317]. Antibody-mediated

blocking of VWF resulted in a 50% reduction of leukocyte recruitment, similar to that observed in VWF-deficient mice[318, 319], with similar effects of VWF seen when injecting autologous blood provoked an inflammatory response[320]. This has been confirmed by the vast majority of animal models[316, 319, 320]. Also reported in animal models and vascular disease patients is an accumulation of extravascular VWF though whether this is plasma VWF or that released directly from endothelial cells on the basal subendothelial matrix remains unclear[318, 320, 321]. The fact that VWF is able to directly interact with β2-integrins expressed on blood-borne leukocytes suggests that VWF could play a chemo-attractant role[322, 323]. Further studies suggested that β2-integrins bind to two fragments in the D'D3 and the A domain region. PSGL-1, a second leukocyte receptor, was identified as binding to the A1 domain alongside the GpIbα binding site[323] (explained by the structural similarities between GpIbα and PSGL-1[324]). Binding of VWF to PSGL-1 occurs when VWF is in its active conformation, as occurs with GpIbα, increasing its inflammatory activity and confirmed by an increased presence of haemoglobin in VWD Type 2B mice (overactive GpIbα binding)[325]. Direct binding between newly released VWF and GpIbα has also been shown to induce increased leukocyte rolling, however, it is unclear if this is by a secondary mechanism whereby this interaction induces the release of other inflammatory stimuli from endothelial cells or by activation of platelets themselves[326, 327].

1.6.2. Angiogenesis

Recent research has emerged showing the distinct overlap of haemostasis, the cessation of blood loss, and angiogenesis, the formation of new blood vessels, with a number of haemostatic proteins implicated in the process of angiogenesis[328-330]. VWF has been shown to be a factor in angiogenesis regulation as VWF-deficient mice and human umbilical vein endothelial cells with inhibited VWF expression display increased levels of angiogenesis[299]. Increased angiogenesis is associated with a number of pathologies including tumours, diabetes and arteriosclerosis[331]. The regulation of angiogenesis by VWF has been suggested to occur by both extracellular and intracellular pathways, implicating endothelial cell surface integrin $\alpha_v\beta_3$ and the vascular GF Ang-2, respectively[332, 333].

The integrin $\alpha_v\beta_3$ has increased expression in proliferating vascular endothelial cells. This surface integrin has been shown to have both pro- and anti-angiogenic depending on a variety of receptors and ligands[332, 334]. VWF binds to $\alpha_v\beta_3$ on the endothelial cell surface which controls the expression of $\alpha_v\beta_3$ expression and prevents its internalisation. This binding occurs under fluid shear stress by the RGD motif found in the VWF C4 domain which binds specifically to the integrin β_3 region[237, 335]. The binding of VWF to integrin $\alpha_v\beta_3$ has an indirect impact on downstream signalling caused by the binding of $\alpha_v\beta_3$ to vascular endothelial GF receptor-2

(VEGFR2)[336]. In normal circumstances, this interaction is essential for the formation of new endothelial cells and angiogenesis in tumours and tissue regeneration[332, 337]. A lack of endothelial β_3 has been shown to result in increased sensitivity of VEGFR-2 signalling and the formation of immature, fragile blood vessels. Whilst VWF has been correlated with decreasing VEGFR-2 sensitivity, it is unclear whether this is by a direct mechanism or, more likely, an indirect mechanism where binding to $\alpha_v\beta_3$ prevents internalisation of the integrin and maintains sufficient levels of endothelial β_3[299, 332]. The fact that patients deficient in high molecular weight VWF multimers display vascular malformations suggests that these are critical for regulation of angiogenesis, whether by recruiting platelets as angiogenesis regulators or interacting with particular circulating ligands that enhances the affinity for their receptor[332]. The impact of VWF deficiency has been reported in a multitude of studies with angiogenesis and vascular density increased in VWF-deficient mice, both in general[299] and in the brain in response to hypoxia[338]. Furthermore, disruption of $\alpha_v\beta_3$ or VWF expression resulted in vascular smooth muscle cell expression, a marker of vascular maturation[339]. The attempts at targeting VWF, $\alpha_v\beta_3$ and VEGFR-2 signalling to combat tumour progression highlight the complexity and delicate balance of the pathway.

The intracellular pathway involves the Angiopoietins/Tie-2 pathway through the interaction of VWF with Ang-2. VWF is stored alongside a number of co-residents in WPBs, including Ang-2 which is a ligand of the receptor tyrosine kinase Tie-2[340]. Binding of released Ang-2 leads to activation of the Angiopoietins/Tie-2 pathway which regulates vascular haemostasis and angiogenesis[332]. With the formation of WPBs and Ang-2 storage relying upon the expression of VWF, a lack of WPB formation due to VWF-deficiency increases angipoietin-2 secretion, possibly leading to increased activation of this pathway and subsequent angiogenesis[299]. The fact that VWF has been shown to remain bound to Ang-2 following secretion at the VWF A1 domain raises a possibility that VWF localises Ang-2 to the Tie-2 receptor or Ang-2 modulates the role of VWF in angiogenesis[340].

1.6.3. Tumour metastasis

Endothelial cells lie at an essential interface in the case of cancer progression through metastasis, the spread of cancer cells to a secondary site. Separating the vasculature from the surrounding tissues, tumour cells understandably target endothelial cells in this process. Whilst a connection between coagulation and cancer has been known for more than 150 years with venous thromboembolism being a significant complication in cancer patients, research has recently indicated that cancer progression can be affected by coagulation pathways. This extends to cancer cells interacting with platelets and coagulation factors, leading to tumour

growth, angiogenesis and metastasis[341, 342]. Platelets have been shown to both protect circulating tumour cells from the body's defence mechanism and facilitate adhesion to and across the endothelium[343]. With VWF being heavily involved in platelet adhesion, it likely plays a role in mediating these effects. Tumour cells have also been shown to express pseudo-GpIbα receptors on their surface, binding to VWF directly[344]. The fact that platelet deficiency and anti-VWF treatment have both been shown to reduce tumour metastasis, their importance in metastasis is clearly evident[345, 346].

Elevated VWF levels have been reported in patients with a variety of tumour types, both haematological and solid tumours, with high levels of plasma VWF correlating with poorer survival[347, 348]. Tumour cells activate endothelial cells through a variety of agonists including cytokines and matrix metalloproteinases, releasing the WPB contents into circulation and resulting in elevated plasma VWF levels[341]. This mechanism was shown to be prolific in aggressive tumours with some inducing endothelial cell activation to the same degree as histamine, one of the most potent activators[349]. Along with elevated release of UL VWF, studies have shown that melanoma cells downregulate ADAMTS13 activity, increasing the half-life of the secreted UL VWF multimers[346].

Accumulation of VWF has also been shown in the tumour stroma and microenvironment, though its mechanism remains poorly understood. Some studies suggest that tumour cells induce increased VWF expression through binding to the VWF promoter region[350], whilst others implicate the VWF A3 domain in binding to exposed collagen associated with solid tumours[351]. This accumulation may influence immune cell recruitment and inflammation[341]. Some cancer cells also acquire the ability to express VWF[352, 353] and store multimers in pseudo-WPBs[341]. The ability of tumour cells to express VWF has been shown to improve cell adhesion to endothelial cells and extravasation across the endothelial barrier[354]. Consistent with this observation, inhibition of tumour-derived VWF expression has been shown to reduce tumour migration[352]. The structural difference between endothelial- and tumour-derived VWF, if any, has yet to be determined[341].

As discussed previously, VWF has a significant impact on inflammation and angiogenesis, two processes critically important in tumour metastasis. With VWF acting as an adhesive receptor on activated endothelial cells to recruit leukocytes, tumour cells adopt similar pathways which can in turn contribute to tumour metastasis by extravasation across the endothelium. Similarly, increased intra-luminal VWF has been shown to have a pro-angiogenic effect, causing increased tumour vessel density[341].

1.7. Von Willebrand Disease

When circulating levels of VWF are lower than normal or there is a defect in the protein, the resulting pathology is known as VWD. This is the most common inherited bleeding disorder in the world and is divided into Types 1-3, with Type 2 subdivided into four subtypes (Table 1.1). Types 1 and 3 are quantitative defects whereas Type 2 is a qualitative defect. In contrast to haemophilia, another clotting disorder that affects clotting factors in the coagulation cascade, VWD is more widespread and generally less severe[355]. Diagnosis of VWD is performed by a multitude of methods including GpIbα-, platelet- and FVIII-binding assays, along with a discussion regarding family bleeding history[356]. A common treatment for VWD is the use of DDAVP, a secretagogue that stimulates the secretion of VWF from storage granules. This treatment has varying levels of success depending on the VWD type of a patient[357]. Patients not suitable for DDAVP treatment may instead be suitable for VWF replacement therapy[356]. VWD-causing mutations are reported in the international database at the University of Sheffield[358] (http://vwf.group.shef.ac.uk/).

Type	Description
1	Partial quantitative defect
2	Qualitative defect
2A	Lacking high molecular weight multimers causing impaired primary haemostasis
2B	Gain-of-function in binding to platelet GpIbα, causes spontaneous platelet binding and increased breakdown by ADAMTS13
2M	Loss-of-function in binding to platelet GpIbα causing impaired primary haemostasis
2N	Impaired binding to coagulation FVIII causing it to be rapidly broken down
3	Almost complete VWF deficiency

Table 1.1: Classification of the VWD types and subtypes

1.7.1. Type 2 VWD

This is a qualitative defect which is divided into four subtypes, 2A, 2B, 2M and 2N. Whilst displaying normal plasma VWF levels, patients exhibiting Type 2 VWD exhibit structural and functional defects in VWF[355]. Although DDAVP would not be expected to provide a haemostatic benefit to Type 2 VWD patients due to the qualitative defects of the stored protein, some subtypes display a clinical response.

Type 2A VWD is characterised by an abnormal multimer pattern where patients lack high molecular weight multimers[359]. This causes reduced levels of platelet binding due to the fact

that these multimers are the most efficient in primary haemostasis. After Type 1, Type 2A VWD is the second most common type, accounting for 10-15% of VWD cases[360]. It is caused by mutations in a range of domains including the D2, D3, A1, A2 and CK domains. Of the 149 reported mutations in the VWF mutation database, approximately 90% of those are missense mutations with those in D2 being recessive mutations and those in the other four domains being dominant[361, 362]. Mutations in the D2, D3 and CK domain, known as Group I mutations, inhibit dimerisation and multimerisation which, therefore, prevents high molecular weight multimers being synthesised[363-366]. These mutations result in intracellular retention and loss of regulated storage[367]. Those mutations in the A1 and A2 domains, Group II mutations, increase the susceptibility to cleavage by ADAMTS13[368]. Therefore, whilst they produce UL VWF multimers, these are rapidly cleaved upon secretion into the vasculature[369]. Group I mutations which prevent multimer formation exhibit a more severe bleeding phenotype though also respond better to treatment with DDAVP[362]. Whilst some variants fit clearly into Groups I or II, others display and overlap of defective intracellular and extracellular processing[367].

Whilst Type 2A exhibits reduced platelet binding, Type 2B VWD is characterised by increased VWF binding to platelets. This is due to a dominant gain-of-function mutation in the VWF A1 domain that causes increased binding to platelet GpIbα[362]. This also results in reduced levels of high molecular weight VWF multimers in plasma due to proteolysis by ADAMTS13 following spontaneous platelet binding[359, 370, 371]. Patients can also exhibit further complications such as thrombocytopenia and platelet agglutination[362]. There have been 74 recorded mutations in the VWF mutation database causing Type 2B VWD with these only detected in the region of exon 28 encoding for the VWF A1 domain. Over 95% of the recorded mutations are missense mutations. Whilst all mutations cause overactive binding to GpIbα, further complications vary depending upon the specific mutation[372, 373].

Conversely to Type 2B, as defect in platelet binding is classified as Type 2M VWD. This is a dominant loss-of-function mutations typically found the platelet GpIbα binding region. Some mutations are also reported in the VWF A3 domain which affect collagen binding, however, these patients respond better to treatment with DDAVP than those with mutations in the GpIbα binding site[362]. There have been 53 reported mutations causing Type 2M VWD in the VWF mutation database with 94% of those being missense mutations and the remainder being deletions. Whilst relatively low numbers have been reported, a study found that Type 2M VWD is misidentified in around 70% of cases[374]. Unlike other forms of Type 2 VWD, Type 2M patients do not display a loss of high molecular weight VWF multimers[375].

The final subtype of Type 2 VWD is Type 2N, characterised by impaired FVIII binding in circulation. This results in rapid breakdown of FVIII in circulation and reduced levels, however, in contrast to Haemophilia A, is not caused by impaired FVIII synthesis[362, 376]. A total of 92 recessive mutations have been reported across the D' and D3 domains, altering FVIII binding affinity. Of those reported, 95% are missense mutations with the remainder caused by deletions. The severity of the phenotype of dependent upon the mutation and their effect on the FVIII binding affinity, with the most severe mutations (E787K, T791M and R816W) being found in the primary FVIII binding site in the D' domain[357, 377].

1.7.2. Type 3 VWD

A total absence or very low levels of circulating VWF is known as Type 3 VWD and accounts for 1% of VWD cases. This is caused by a wide range of mutations throughout the whole protein with missense, nonsense, insertion and deletion mutations all leading to Type 3 VWD. The most common Type 3 VWD-causing mutation is a nonsense mutation, the most common of which is R1659X occurring in the VWF A2 domain[362]. A nonsense mutation would cause endless issues in preventing dimerisation and multimerisation, storage in WPBs and leading to retention within the endothelial cell. The second most common mutation type leading to Type 3 VWD is a missense mutation, some of which are of cysteine residues which may cause issues with protein folding and multimerisation[378]. All 316 mutations leading to Type 3 VWD are recessive with a patient being homozygous or heterozygous with one null allele. As would be expected with Type 3 VWD patients with their issues in storing VWF multimers, treatment with DDAVP is not suitable though can be treated with VWF replacement therapy[356].

1.7.3. Type 1 VWD

A slight quantitative deficiency of circulating VWF, usually in the range of 20-40%, is classed as Type 1 VWD[379]. This type, where patients display normal multimer patterns, is the most common type of VWD accounting for 60-70% of cases and is typically the least severe[380]. In more severe cases of VWD, genetic mutations are the dominant causation, however, Type 1 VWD-causing mutations have been shown to not be limited to the VWF gene. This has been shown with mutations in a number of clearance receptors associated with Type 1 VWD[381]. Mutations of the VWF gene that do arise from Type 1 VWD largely cause issues regarding intracellular transport and lead to intracellular retention[355]. Sequence variations in the VWF gene of Type 1 VWD patients has been shown to be relatively widespread, with a higher number of genetic variation shown in patients with a more severe phenotype. Whilst these mutations are well defined, the molecular mechanisms by which they affect the overall

structure are as yet poorly understood. This is the case with the mutation that has been suggested as the most common VWD Type 1 mutation, Y1584C in the VWF A2 domain. The exact mechanism mediated by this change is still debated[382].

Around 80% of Type 1 VWD-causing mutations arise from missense mutations with the remaining 20% made up of promoter, splice site and nonsense mutations, along with insertions and deletions[362]. Studies have indicated that these mutations cause Type 1 VWD in one of two ways. Firstly, it can lead to an increased rate of VWF clearance from circulation. Patients that display this increased level of clearance are placed in a distinct category, Type 1C. Typical of increased VWF clearance, patients display an increased VWF propeptide/plasma VWF antigen ratio[383, 384]. Increased clearance reduces the effectiveness of DDAVP treatment with normal VWF levels lasting less than 4 hours[362]. Mutations that increased VWF clearance include that of R1205 with mutations at this site shown to have enhanced binding to THP-1 macrophages, increasing macrophage-induced VWF clearance[385]. Genetic mutations throughout the VWF gene have been associated with increased VWF clearance, however, the mechanism by which this arises needs further investigation[381].

The second method that causes Type 1 VWD is intracellular retention of the mutant VWF protein. These mutations lead to misfolding and loss of multimeric structure and are often caused by a loss or gain of cysteine residues[362]. Amongst the C domains of VWF, on which this project focuses, the C3 domain displays the most VWD Type 1-causing mutations. A total of nine mutations have been reported, 5 of which are caused by a gain or loss of cysteine residues, leading to misfolding and increased levels of intracellular retention[382, 386]. Other than those caused by a loss or gain of cysteine, one is a serine to proline mutation, however, this is in combination with a nonsense mutation earlier in the VWF genetic sequence[386]. A nonsense mutation in C3 was also reported as causing Type 1 VWD, surprisingly considering the subsequent lack of a C-terminus and dimerisation ability[387]. Other mutations occur by insertions of a single nucleotide. Interestingly, all patients which display a loss or gain of cysteine residues display an abnormal multimer pattern though a definitive reason for this remains undetermined[382, 386]. This is unlikely due to increased VWF clearance susceptibility with the C domain region seemingly not having an influence in this regard. It may simply be due to misfolding and a higher rate of intracellular retention of UL and high molecular weight multimers. Another explanation may be impaired UL VWF multimer formation and storage with the gain or loss of cysteines affecting the “stalk” of the bouquet structure in multimerisation, in turn reducing high molecular weight multimers in circulation. Whilst the “stalk” structure has been observed by microscopy[190], a structure on the molecular level has

yet to be determined. Further studies may provide an insight into how the gain or loss of cysteines in this case may affect its arrangement and, subsequently, storage of UL VWF multimers.

1.8. Structure determination using NMR spectroscopy

The discovery of NMR dates back to Isidor Isaac Rabi in 1944, with Felix Boch and Edward Mills Purcell first demonstrating NMR in condensed matter in 1945. The field of NMR has developed incomprehensibly since then, now used to investigate the structure of organic molecules, molecular dynamics (MD) and in advanced medical imaging techniques (magnetic resonance imaging). Its use in structural biology is highly popular due to its accuracy, particularly in small, flexible proteins. Other methods of structural determination include X-ray crystallography and cryogenic EM (cryo-EM). Both methods have their advantages, with X-ray crystallography being a relatively cheap and simple method which can yield structures with high atomic resolution. It does not, however, provide any insight into protein dynamics in solution as is the case with NMR. Furthermore, X-ray crystallography requires the protein to be crystallised, something which has not been possible in other VWF domains. Cryo-EM, meanwhile, is particularly useful as it requires a small amount of sample and the rapid freeze treatment of the sample maintains its closer-to-native state compared to X-ray crystallography. It does, however, have relatively low resolution and also lacks the capability to provide insight into protein dynamics. Whilst NMR requires high sample purity which presents challenges in sample preparation, its high resolution and capacity to investigate protein dynamics make it the most suitable method to determine the structure of the VWF C3 domain.

NMR requires the use of certain isotopes, in this project ^{1}H, ^{13}C and ^{15}N, each of which have a spin-½ nucleus due to their odd number of nucleons. This results in each nucleus exhibiting two energy states which, once exposed to an external magnetic field, split into a low and high energy level, α and β, respectively.

1.8.1. Chemical shift

With nuclei remaining in their energy state at equilibrium and oscillation between the two being rare, application of a radio-frequency energy equal to the difference between the two energy states will result in the oscillation between the two states, known as the Larmor frequency. This is followed by relaxation to their ground state, known as free induction decay. This measurement, acquired in the time domain, can be transformed into the frequency domain, a process known as Fourier transform (Figure 1.18). The chemical shift is defined by

the frequency of a resonance with reference to a standard compound, usually sodium trimethylsilylpropanesulfonate (DSS). The chemical shift, δ, in parts per million (ppm) of a resonance in a sample is calculated by the following equation:

$$\delta_{sample} = \left(\frac{\nu_{sample} - \nu_{reference}}{\nu_{reference}} \right) x\ 10^6$$

where *v* is the absolute frequency of the sample resonance or reference compound. Chemical shift is dependent upon the type of nucleus, such as a ^{1}H, ^{13}C or ^{15}N and the surrounding chemical environment. Whilst the Larmor frequency is proportional to the external magnetic field, the chemical shift is normalised using the equation detailed above. Chemical shifts, therefore, can provide an insight into an amino acid's structural profile whether it is involved in secondary structure, resides in a hydrophobic core or is freely exposed.

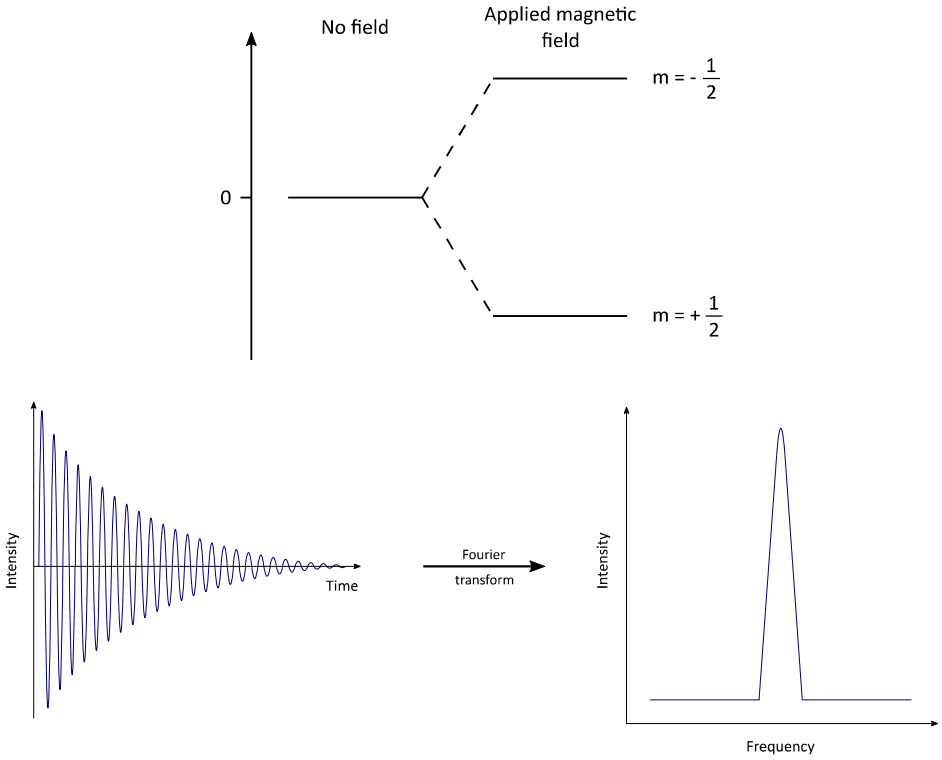


Figure 1.18: Energy levels, free induction decay and Fourier transform Displayed above is the energy level splitting of a spin ½ nucleus upon the application of a magnetic field. Denoted next to each energy level is its magnetic quantum number, *m*. Below is a free induction decay which shows the NMR signal generated by applying a magnetic pulse. This signal in the time domain is transformed into the frequency domain by a Fourier transformation.

The effect of the surrounding chemical environment is what makes the measurement of chemical shift particularly useful in determining protein structures with peaks well dispersed and distinguishable from each other in folded proteins. This allows chemical shift data to be used to predict regions of secondary structure, along with tertiary structure when used in

combination with a number of parameters, such as inter-proton distances (NOEs). A distinct advantage of using chemical shifts is that they are readily and accurately measurable.

1.8.2. J-Couplings

An essential element to a number of NMR experiments is the concept of J-couplings, otherwise known as spin-spin or scalar coupling. These are through-bond couplings which arise when a spin of one nucleus polarises the spins of the surrounding electrons. These polarised electrons will, in turn, polarise neighbouring nuclei. A J-coupling will occur between two spins which are chemically bound and only a few bonds apart. It is field-independent, although, due to J-couplings being a through-bond interaction, their magnitude rapidly decreases as the number of intervening bonds increase. This phenomenon has become essential in determining protein structures, regularly used in NMR experiments that allow for a sequential protein assignment.

1.8.3. NMR experiments and chemical shift assignment

There are various NMR experiment types used in the chemical shift assignment of a protein. One of the simplest NMR experiments is the ^{15}N-HSQC (**H**eteronuclear **S**ingle **Q**uantum **C**oherence), often referred to as the protein 'fingerprint', which plots the chemical shift of the amide nitrogen against that of the attached proton and presents a single peak for each amino acid. This is true for all except proline, which does not contain an amide proton when in sequence, whilst cross-peaks are also displayed for the side chain asparagine ($^{1}H_{\delta}$-$^{15}N_{\delta}$), glutamine ($^{1}H_{\varepsilon}$-$^{15}N_{\varepsilon}$) and tryptophan ($^{1}H_{\varepsilon}$-$^{15}N_{\varepsilon}$). These are, however, easily distinguishable from the backbone amide cross-peaks, with the $^{15}NH_2$ groups of asparagine and glutamine visible in the ^{1}H dimension of 6.5-8 ppm and around 95 ppm in the ^{15}N dimension. These peaks can be removed using a decoupled ^{15}N-HSQC, however, this would not be necessary due to them being easily distinguishable from the backbone amide groups. The tryptophan side chain $^{1}H_{\varepsilon}$-$^{15}N_{\varepsilon}$ group is often well separated from the backbone amide cross-peaks with the average chemical shift of the $^{1}H_{\varepsilon 1}$ of tryptophan being 10.09 ppm. Cross-peaks from certain arginine side chain nuclei ($^{1}H_{\varepsilon}$-$^{15}N_{\varepsilon}$) only tend to be observable at an acidic pH, below 6-7, whilst other arginine side chain nuclei ($^{1}H_{\eta}$-$^{15}N_{\eta}$), along with histidine ($^{1}H_{\delta 1}$-$^{15}N_{\delta 1}$, $^{1}H_{\varepsilon 2}$-$^{15}N_{\varepsilon 2}$) and lysine ($^{1}H_{\zeta}$-$^{15}N_{\zeta}$) side chain nuclei are only visible when in the protein core, involved in hydrogen bonds or form salt bridges.

With each peak in a ^{15}N-HSQC corresponding to a single amide group, assignment of the chemical shift of each backbone amide group is the essential first step to a full chemical shift assignment and subsequent structure calculation. This requires a series of triple resonance

experiments which allow a sequential backbone amide assignment to be performed. The four triple resonance experiments primarily used were the HNCO, HN(CA)CO, HNCACB and HN(CO)CACB. The names of each experiment refer to the chemical shift of which nuclei they display whereas those nuclei in brackets are involved in magnetisation transfer but not chemical shift evolution. For example, the HNCO correlates the backbone amide ^{1}H against the amide ^{15}N of residue *i* in two of the three dimensions. The third dimension correlates with the $^{13}C_O$ chemical shift of the preceding residue, *i-1*. The HN(CA)CO, however, due to the magnetisation transfer at $^{13}C_\alpha$, displays peaks correlating to the $^{13}C_O$ chemical shift of both residues *i* and *i-1*. Similarly, the HNCACB correlates the chemical shift of the backbone amide ^{1}H and ^{15}N of residue *i* in two of the three dimensions against the $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts of both residues *i* and *i-1* in the third dimension. The HN(CO)CACB, due to the magnetisation transfer through $^{13}C_O$, solely displays peaks for the $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shift residue *i-1* in the third dimension. Analysis of these experiments allows not only an assignment of the backbone amide group but also the $^{13}C_O$, $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts.

A further experiment used in the backbone amide assignment is the CC(CO)NH-TOCSY (**TO**tal **C**orrelation **S**pectroscop**Y**) which correlates the backbone amide ^{1}H and ^{15}N chemical shifts of residue *i* with the side chain ^{13}C nuclei of residue *i-1*. This is particularly useful for determining the residue type for amino acids which display unique side chain ^{13}C chemical shift patterns such as arginine or isoleucine. Once the backbone amide assignment is completed, the chemical shift of the side chain nuclei must be determined. This assignment requires the HCC(CO)NH-TOCSY and HCCH-TOCSY experiments, along with the CC(CO)NH-TOCSY, to plot the chemical shift of the side chain ^{1}H and ^{13}C nuclei on a 2D ^{13}C-HSQC. Similarly to the ^{15}N-HSQC, a ^{13}C-HSQC plots the chemical shifts of a ^{1}H and ^{13}C group which exhibit a one-bond coupling. The HCC(CO)NH-TOCSY is of a similar format to the CC(CO)NH-TOCSY, however, rather than plotting the chemical shift of the side chain ^{13}C nuclei of residue *i-1*, it plots the chemical shift of the side chain ^{1}H nuclei of residue *i-1*. The HCCH-TOCSY plots the chemical shift of an aliphatic ^{1}H nucleus with a cross-peak to all ^{1}H nuclei in the aliphatic side chain, provided both ^{1}H nuclei are bound to a ^{13}C nucleus. Each of these experiments for the side chain assignment were performed as a TOCSY rather than a COSY (**CO**rrelation **S**pectroscop**Y**). This is a benefit in, for example, the HCCH-TOCSY where at position $^{1}H_\alpha{}^{13}C_\alpha$ a TOCSY will theoretically present a cross-peak for each ^{1}H in the side chain ($^{1}H_\beta$, $^{1}H_\gamma$, $^{1}H_\delta$ etc). A COSY, on the other hand, will only present a cross-peak for the neighbouring nuclei (i.e. position $^{1}H_\alpha{}^{13}C_\alpha$ only presents a cross-peak for $^{1}H_\alpha$ and $^{1}H_\beta$). A further experiment used is the DOSY (**D**iffusion **O**rdered **S**pectroscop**Y**)

which separates NMR signals according to their diffusion coefficient. Such an experiment differentiates peaks which originate from either the same or different populations.

The aliphatic chemical shift data contributes to an accurate structure calculation, however, this accuracy can be improved by the stereospecific assignment of prochiral groups of protons. This assignment allows more detailed analysis of side chain interactions which may otherwise not be possible[388]. This can be achieved by expressing a protein sample using 10% $^{13}C_6$-D-glucose and 90% $^{12}C_6$-$^{1}H_{12}$-glucose, known as fractional labelling[389]. Discounting the natural ^{13}C abundance of $^{12}C_6$-$^{1}H_{12}$-glucose, there is a 10% chance of any given carbon nucleus being ^{13}C-labelled, therefore, there is a 1% chance of two neighbouring carbon nuclei being ^{13}C-labelled. This is true unless these originate from the same carbon source, in which case if one is ^{13}C-labelled then the other will also be.

The principal of using a 10% ^{13}C-labelling method for a stereospecific assignment relies upon the stereospecific biosynthesis of valine and leucine. Our understanding of the stereochemistry of substrates and reaction pathways allows non-random labelling patterns to be exploited to yield a stereospecific NMR assignment. In the case of the Pro-R methyl group ($^{13}C_{\gamma 1}$ in valine and $^{13}C_{\delta 1}$ in leucine), the neighbouring carbon nucleus ($^{13}C_{\beta}$ in valine and $^{13}C_{\gamma}$ in leucine) originates from the same carbon source, however, that of the Pro-S methyl group ($^{13}C_{\gamma 2}$ in valine and $^{13}C_{\delta 2}$ in leucine) originates from a different carbon source. Therefore, if the Pro-R methyl group is ^{13}C-labelled, as it will be in 10% of cases, there is a 100% chance the neighbouring carbon is also ^{13}C-labelled, however, if the Pro-S methyl group is ^{13}C-labelled, there is a 10% chance that the neighbouring carbon nucleus is also ^{13}C-labelled (Figure 1.19).

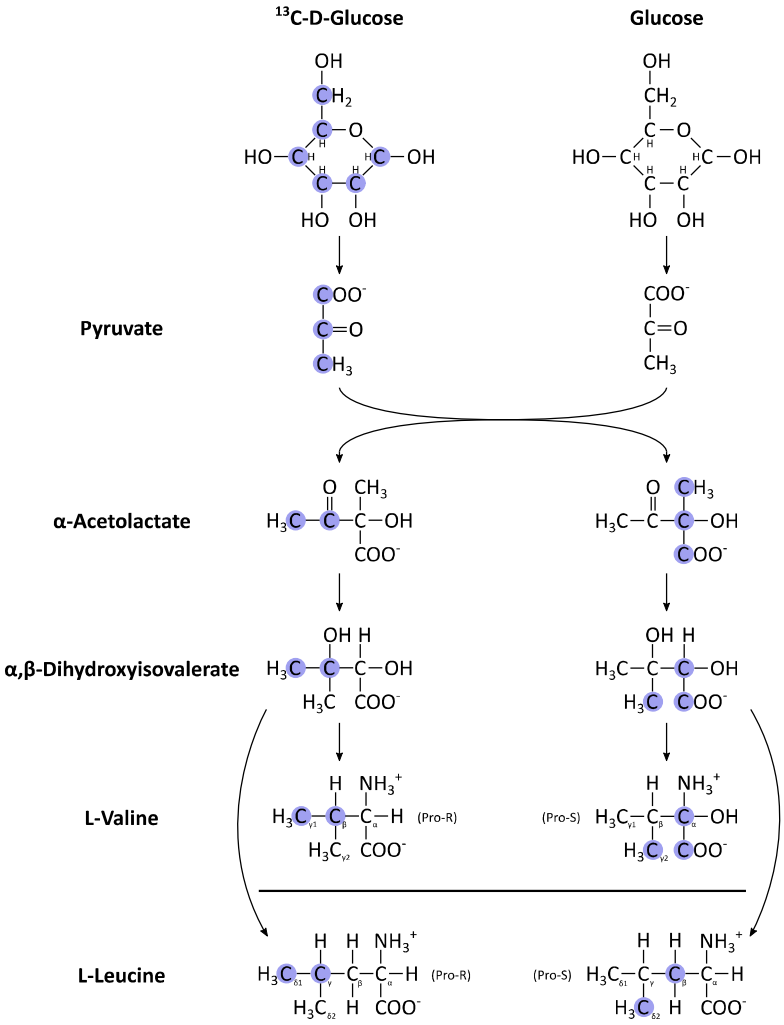


Figure 1.19: Biosynthesis of valine and leucine residues using fractional ^{13}C-labelling Displayed is the biosynthesis sequence of valine and leucine residues using a mix of 10% $^{13}C_6$-D-glucose and 90% $^{12}C_6$-$^{1}H_{12}$-glucose, whereas the migration of methyl groups is stereoselective. Carbon nuclei highlighted in blue indicate a ^{13}C-labelled nucleus while those unhighlighted indicate a natural abundance ^{13}C nucleus[390].

Peaks corresponding to Pro-R and Pro-S ^{13}C nuclei reside in opposing phases, positive and negative, with which being which depending on the phasing of the peaks during processing of the spectrum. Pro-R methyl groups reside in one phase due to an even number of bound ^{13}C nuclei ($^{13}C_{\gamma 1}$-$^{13}C_{\beta}$ in valine and $^{13}C_{\delta 1}$-$^{13}C_{\gamma}$ in leucine), conversely, Pro-S methyl groups will be in the opposite phase with an odd number of ^{13}C nuclei directly bound ($^{13}C_{\gamma 2}$ in valine and $^{13}C_{\delta 2}$ in leucine). This arises from the ^{13}C-^{13}C coupling between the ^{13}C-labelled Pro-R methyl carbon with the ^{13}C-labelled connected carbon nucleus. This coupling is lacking in the case of the ^{13}C-labelled Pro-S methyl carbon nucleus with its adjoining carbon being ^{13}C-labelled in only 10% of cases.

1.8.4. Nuclear Overhauser effect

In contrast to J-couplings which are through-bond interactions, the interaction of nuclear spins by through-space magnetic interactions is known as the nuclear Overhauser effect (NOE). It works on the premise that each nuclear spin generates its own magnetic field when exposed to an external magnetic field. Each nuclear spin does not relax independent of its neighbouring spins, rather the magnetic field of a neighbouring dipole, *spin S*, will alter the spin population of its neighbour, *spin I*, provided the two are in close proximity (< 5 Å). This interaction occurs between bonded nuclei and those close in space, with the dipole-dipole coupling resulting in a NOE[391].

A NOE arises when a transfer of magnetisation occurs between two protons, *I* and *S*, through dipolar mechanisms. NOEs between two ^{1}H nuclei can be observed as cross-peaks in a series of spectra known as NOESY (**N**uclear **O**verhauser **E**ffect **S**pectroscop**Y**) spectra. A NOESY spectrum exhibits two ^{1}H dimensions along with either ^{13}C or ^{15}N in the third dimension in the case of a ^{13}C- or ^{15}N-edited NOESY, respectively. In the case of the ^{13}C-edited NOESY, a diagonal peak displays the chemical shift of a ^{13}C-bound ^{1}H nucleus with an identical chemical shift in both the direct and indirect ^{1}H dimensions. In the ^{13}C dimension is the chemical shift of its bound ^{13}C nucleus. A series of cross-peaks are observed in line with this diagonal peak with different chemical shifts in the indirect ^{1}H dimensions. The chemical shift of each of these cross-peaks in the indirect ^{1}H dimension corresponds to the chemical shift of a neighbouring ^{1}H nucleus. These nuclei are close in space and, therefore, provide insight into a protein's secondary and tertiary structure. Not only can NOESY spectra allow the assignment of two nuclei that are close in space, they can also be used to assign the chemical shifts of aromatic nuclei. The qualities of NOESY spectra and the information the fact that they indicate nuclei which are close in space but not necessarily bound have made them an essential tool in a structure calculation.

1.9. Project aim

The sole C domain of VWF to have its structure solved is the C4 domain, containing the functional $\alpha_{IIb}\beta_3$ binding site, however, as the C4 domain is termed a VWFC-like domain, the structure of a VWFC domain of VWF itself remains to be characterised. Whilst the C3 domain does not contain any functional binding sites, it is classed as a VWFC domain and its importance in the expression and secretion of a full-length VWF multimer has been shown where a VWF deletion variant lacking the C3 domain failed to be secreted[392]. Though multiple mutations in the C3 domain that lead to Type 1 VWD have been identified, it is impossible to understand their impact on a structural level without having a starting point, in this case, the structure of the WT C3 domain. It is unlikely that these mutations have a significant effect on the primary functionality of VWF, recruitment of platelets to a site of vessel injury. Certain cysteine mutations may affect lateral self-association, however, studies have yet to show that the loss of a single cysteine can significantly affect VWF functionality. With multiple cysteines involved in this process, the workload of self-association can likely be managed by alternative cysteines. The most plausible explanation for cysteine-related mutations causing Type 1 VWD is misfolding, leading to intracellular retention due to a highly significant structural impact. It would be interesting to determine which disulphide bonds are particularly essential in the folding of the WT C3 domain and whether those mutated in the cases of Type 1 VWD appear to severely affect domain folding. In this project, we will determine the structure of the VWF C3 domain, providing a starting point to progress onto the structural effects of these mutations.

This work can provide a basis to understand other emerging areas of significance, such as free thiols. The impact of these structural changes in circulation can be determined and how they affect domain and multimer flexibility, aiding VWF functionality. Of particular significance are the C2431-C2453 and C2451-C2468 disulphide bonds with their importance in lateral self-association having been demonstrated[274]. Interactions between the C domains can also be investigated and how they impact the function of VWF multimers under conditions of shear stress. Our understanding of the bouquet “stalk” structure can be improved, identifying the structural differences and rearrangement between a C3 domain stored in a WPB and in circulation, gaining an insight into structural rearrangement upon VWF secretion. All these areas require a starting point, the structure of the WT C3 domain, in order to gain a comprehensive understanding. We used NMR spectroscopy to determine the structure of the VWF C3 domain.

2. Results: Expression, purification and NMR backbone assignment of VWF C3

Identification of the domain boundaries of the VWF C3 domain and the selection of an appropriate template plasmid was vitally important to synthesise a pure and properly folded C3 sample to be used for NMR spectroscopy. The C3 domain was identified as a VWFC domain by sequence homology where the C domain region was found to contain three VWFC domains, C1, C3 and C5, separated by three intervening regions with VWFC-like properties in length and sequence, C2, C4 and C6[37]. Separating the VWF C3 domain from the C2 domain is a 26 residue serine and threonine-rich linker along with four internally disulphide-bonded cysteines[130]. The C3 domain is flanked at its C-terminus by the C4 domain, in the absence of a linker. The C3 domain spans 68 residues from position 2429 to 2496 of full length WT VWF.

Figure 2.1: VWF C3 sequence and predicted disulphide bonds The VWF C3 sequence is shown with the ten conserved cysteine residues highlighted in blue. Disulphide bonds predicted by sequence analysis and comparison to solved VWFC domain structures[37, 47] are shown with the connecting black lines.

The VWF C3 domain contains ten cysteine residues (15% of residues), all of which are expected to be involved in disulphide bonds (Figure 2.1), therefore, using a template plasmid and maintaining oxidising conditions throughout the expression and purification would be vitally important. Whilst expressing recombinant protein using *Escherichia coli* (*E. coli*) is beneficial to producing large quantities of protein with its unparalleled fast growth kinetics using a rich, complex media from inexpensive and readily available compounds, WT *E. coli* has extensive redox pathways that prevent disulphide bond formation[393]. The redox pathway in *E. coli* first began to emerge in 1993 with multiple subsequent studies providing a clear picture of this pathway[394-396]. Whilst WT *E. coli* was shown to contain a disulphide bond isomerase C (DsbC), the numerous thiol reductases in the thioredoxin B (TrxB) and glutaredoxin/glutathione (gor) pathways meant that cysteines were maintained in their reduced state[397]. This led to the mutation of these two reductive pathways in the cytoplasm of *E. coli*, however, this meant certain essential proteins could not be reduced into their active state[398]. An *E. coli* strain, sold as Origami by Novagen, was developed where Grx1, a thiol reductase, gained its reducing capability and allowed essential proteins to be activated[396]. Whilst successful in oxidising protein substrates[399], in a number of instances proteins were mis-folded due to inaccurate

oxidation. This finally resulted in the development of SHuffle *E. coli* cells, designed to improve consistency in producing correctly oxidised and folded proteins. Based on a TrxB gor strain, SMG96, it acts by over-expressing DsbC and results in efficient cytoplasmic expression and folding of disulphide-bonded proteins[400]. Due to the heavily disulphide-bonded nature of VWF C3, SHuffle *E. coli* cells were chosen as the most appropriate *E. coli* strain for C3 protein expression.

Thioredoxins are a group of small redox proteins which can catalyse disulphide bond formation, whilst simultaneously reducing non-native disulphide bonds in cytoplasmic proteins[395, 401]. A pET-32b(+) vector was used as a template as it contains an N-terminal 11.8 kDa TrxA tag which mimics the VWF prodomains. These contain thioredoxin-like sequences and it has been shown that vicinal cysteines aid disulphide bond formation. TrxA also aids the correct formation of disulphide bonds. Unfolded proteins are translocated to the periplasm of *E. coli* following synthesis with the cytoplasm not permissive for protein folding. Using a series of Dsb proteins, the *E. coli* cell is able to form disulphide bonds and verify their arrangement. A disulphide-bonded complex is formed between DsbA and its substrate, resulting in a reduction in DsbA's active site cysteines and a disulphide bond forming within the substrate protein. Whilst the DsbA active site cysteines are reoxidised by DsbB, the substrate disulphide bond is checked by DsbC. If the protein is deemed to be misfolded, the disulphide bond will either be reduced by DsbC, giving DsbA a second chance to correctly oxidise the cysteines, or isomerised back to their correct oxidised state by DsbC. DsbC, which is reduced when active, is maintained in its active state by DsbD, which importantly receives its electrons from TrxA[402].

Co-expression of TrxA as an oxidative catalyst has been shown to improve the efficiency of disulphide bond formation. Whilst disulphide bonding continues to occur in the absence of an N-linked TrxA tag, due to the high chance of protein misfolding in a protein with ten cysteine residues, the presence of TrxA is essential for the policing of misfolded protein[399]. TrxA has been shown to be far more efficient than DsbA in promoting folding as it occurs more slowly, allowing intramolecular isomerisation and, therefore, favouring the formation of the correct disulphide bonds[399, 403]. The importance of DsbC in yielding correctly folded proteins has been shown particularly in proteins that require non-consecutive disulphide bond formation, as the C3 domain was expected to be. The reductive capabilities of DsbC is hypothesised to be due to its hydrophobic cleft discriminating against interacting with a correctly folded protein whose hydrophobic residues are sheltered within its core, rather reducing the disulphide bonds of proteins whose hydrophobic residues are exposed due to mis-oxidation[402]. Although some

proteins have been expressed in their reduced form and re-folded in an oxidising buffer[125], when such a method was attempted using the C3 domain, the HSQC spectrum indicated a lack of secondary structure with peaks highly concentrated in the centre of the spectrum. Co-expression of unbound TrxA, however, was not performed and may in hindsight be a possible expression strategy.

Connecting the TrxA tag and VWF C3 protein was a 53 amino acid linker which also comprised a 6 x Histidine (His_6) tag, used for purification with nickel-nitrotriacetic acid (Ni-NTA) affinity chromatography. In order for a pure, properly folded sample of VWF C3 to be obtained for NMR spectroscopy, the TrxA tag and linker would need to be removed. Therefore, a FXa cleavage site (Ile-Glu-Gly-Arg$^{\downarrow}$) was used at the C-terminus of the linker. Selection of a suitable cleavage protein was important, as due to the number of disulphide bonds in VWF C3, the folded protein must remain in oxidising conditions. This would create issues for certain proteases, for example Tobacco etch virus (TEV) protease which requires 1 mM dithiothreitol (DTT) or 0.5 mM tris (2-carboxyethyl) phosphine (TCEP) as a reducing agent, or a redox pair (3 mM reduced glutathione [GSH]/ 0.3 mM oxidised glutathione [GSSG]) for disulphide-bonded proteins. However, the low redox potential of -260 mV would risk disulphide bond rearrangement and a misfolded structure for the expressed protein[404], therefore, alternative cleavage peptides were considered. The FXa cleavage site was placed at the N-terminus of the VWF C3 sequence, without a linker between.

2.1. VWF C3 (FXa) construct form, expression and purification

Uncleaved TrxA-His_6-VWFC3 would be purified by Ni-NTA affinity chromatography with the His_6 tag binding to the column and separating the uncleaved protein from other expressed proteins. The TrxA-His_6 tag would be cleaved with FXa and separated from VWF C3 by loading the sample onto a Ni-NTA column and the TrxA-His_6 tag binding and C3 VWF collected in the flow through. Subsequent cleavage, reverse Ni-NTA chromatography and size exclusion chromatography would be used to obtain a pure VWF C3 sample (Figure 2.2).

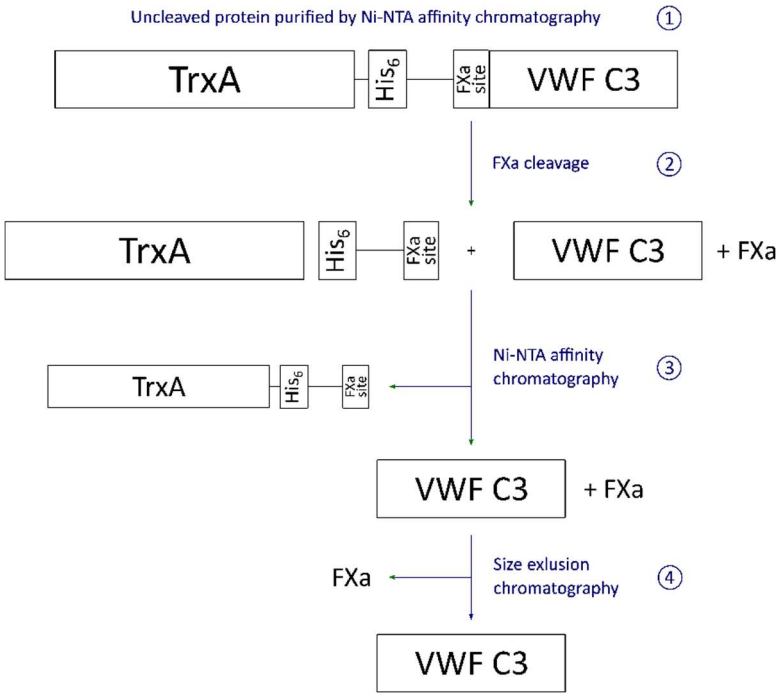


Figure 2.2: VWF C3 purification plan The displayed purification plan is preceded by a Ni-NTA affinity chromatography step (Step 1), which would purify uncleaved TrxA-His_6-VWFC3 from expressed proteins in the *E. coli* cell cytoplasm which lack the His_6 tag. VWF C3 would be cleaved by FXa in optimal conditions which would be determined (Step 2). The TrxA-His_6 tag would be separated from VWF C3 by Ni-NTA affinity chromatography (Step 3), with VWF C3 lacking a His_6 tag and being collected in the flow-through. A pure sample of VWF C3 would then be purified from remaining impurities by size exclusion chromatography (Step 4).

A plasmid obtained from Dr Tom McKinnon encoding the VWF C3 domain was expressed, however once sequenced, it was discovered that following the C3 sequence, there was a 26 amino acid linker followed by a second His_6 tag preceding the termination codon. This gave an expressed protein sequence of TrxA-His_6-VWFC3-His_6 which would provide difficulties during purification. Following FXa cleavage, a reverse Ni-NTA would likely not separate the cleaved TrxA-His_6 fragment from the VWFC3-His_6 protein. Furthermore, the extra 32 amino acids comprising the C-terminal linker and His_6 tag would likely result in issues regarding the NMR assignment with overlapped peaks, the potential for mis-assignment and possible interaction with the C3 domain. A nonsense mutation, therefore, was carried out using polymerase chain reaction (PCR) site-directed mutagenesis to insert a termination codon immediately following the C3 sequence (TGC to TGA) in order to remove the C-terminal linker and His_6 tag from the expressed protein (Figure 2.3).

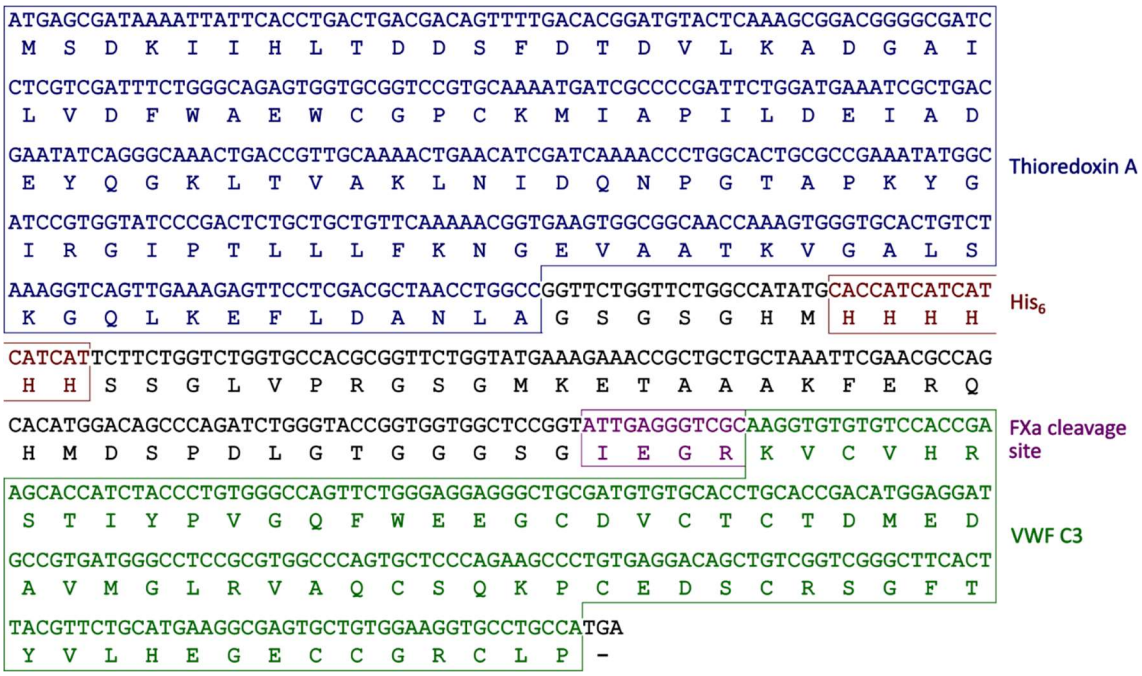


Figure 2.3: Post-nonsense mutation construct sequence Displayed is the DNA sequence and translation product following the nonsense mutation. The product takes the same form, with a 109 amino acid TrxA tag (blue) preceding a His_6 tag (red), separated by a 7 amino acid linker. This is followed by a 36 amino acid linker to a FXa cleavage site (purple) which immediately precedes the 68 amino acid VWF C3 protein (green). This is followed by a termination codon at the C-terminus.

The protein was expressed in M9 minimal media using unlabelled $^{14}NH_4Cl$ and $^{12}C_6$-$^{1}H_{12}$-glucose. Uncleaved TrxA-His_6-VWFC3 was purified over a Ni-NTA column, eluting in a single peak by approximately 180 mM Imidazole. Due to the low concentration of expressed protein and lack of isotopic labelling, this sample was ideal for optimising cleavage conditions, with a time course of 48 hours adopted. Cleavage was carried out using 1 µL FXa/20 µg protein, as suggested by the manufacturer New England BioLabs (NEB), and at 4°C to minimise protein degradation, however, analysis by reducing SDS-PAGE (**s**odium **d**odecyl **s**ulphate-**p**oly**a**crylamide **g**el **e**lectrophoresis) results showed that protein degradation may occur rapidly (Figure 2.4, Panel B). This may, however, have been a one-off case and once optimal cleavage conditions had been identified, the uncleaved sample would be cleaved immediately following purification. Protein stability would be extensively tested further down the line once cleavage and purification had been optimised, testing optimal buffer conditions, pH and temperature to balance protein stability and efficiency in carrying out a series of NMR experiments.

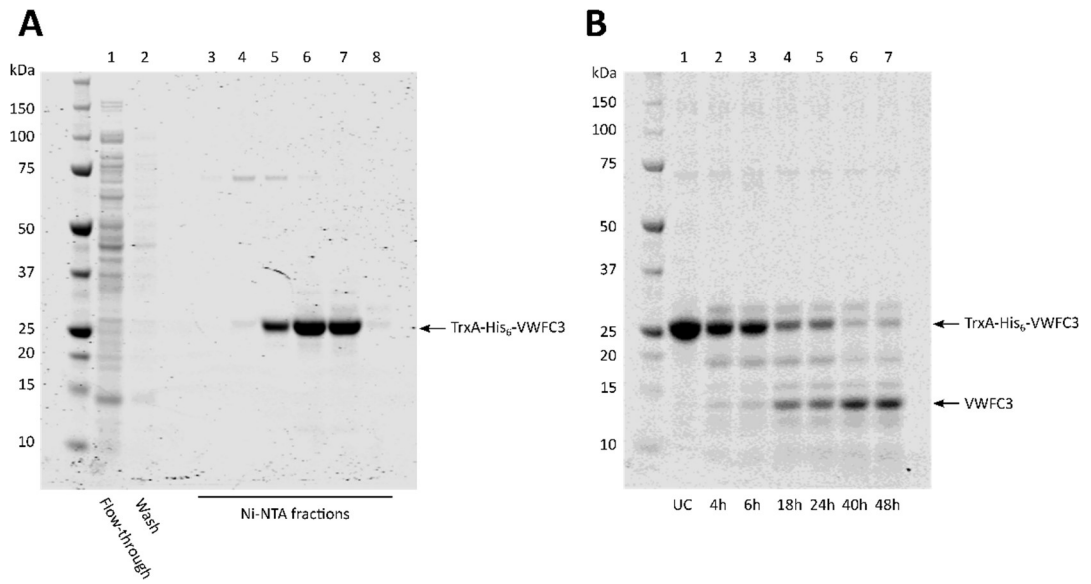


Figure 2.4: Ni-NTA purification of the unlabelled TrxA-His$_6$-VWFC3 and cleavage optimisation Panel A shows the analysis of eluted Ni-NTA fractions in reducing SDS-PAGE, with Lane 1 and 2 showing the unbound material that flowed through and were washed off the column, respectively. Lanes 3-8 display the eluted fractions, with the 25 kDa band in Lanes 4-8 corresponding to the TrxA-His$_6$-VWFC3 protein. Panel B shows the small-scale cleavage of TrxA-His$_6$-VWFC3, with Lane 1 showing the uncleaved protein (UC) and Lanes 2-7 showing cleavage from between 4 hours and 48 hours. The most efficient cleavage was 40 and 48 hours, with these displaying the faintest band at 27 kDa and strongest band at 13 kDa.

Purification of the 24.8 kDa uncleaved TrxA-His$_6$-VWFC3 gave a pure sample with minimal impurities as shown using SDS-PAGE (Figure 2.4, Panel A) with the only potential impurity seen at around 75 kDa, though this may be a TrxA-His$_6$-VWFC3 trimer that had been incompletely reduced. The cleavage over the time course (Figure 2.4, Panel B) suggested that a cleavage time of 40 hours is sufficient, with the band showing the uncleaved sample becoming weakest after 40-48 hours alongside a band becoming noticeably stronger at around 13 kDa. This was deemed too large for the 7.5 kDa VWF C3 fragment but too small for the 17.3 kDa TrxA-His$_6$ fragment, however, the reason for this discrepancy in size became clear later (see Chapter 2.5). In order to minimise the cleavage time to avoid unnecessary protein degradation which may occur during cleavage, 40 hours was deemed sufficient. Whilst a cleavage and purification were attempted with this sample, the low concentration meant that very low UV signal was observed, peaks were relatively indistinguishable and protein concentrations were too low for bands to be observed using SDS-PAGE. Also, due to unlabelled $^{14}NH_4Cl$ and $^{12}C_6$-$^{1}H_{12}$-glucose being used in the expression, analysing the protein by NMR using only ^{15}N natural abundance (0.36%) would be nigh on impossible with such a low protein concentration. This expression did, however, provide a useful insight into efficient cleavage timings which would be used going forward.

2.2. VWF C3 purification and NMR analysis

A ^{15}N-labelled expression was performed in M9 minimal media enriched with $^{15}NH_4Cl$ and unlabelled $^{12}C_6$-$^{1}H_{12}$-glucose. The C3 protein was cleaved and purified, as was the cleaved TrxA-His_6 tag for further analysis.

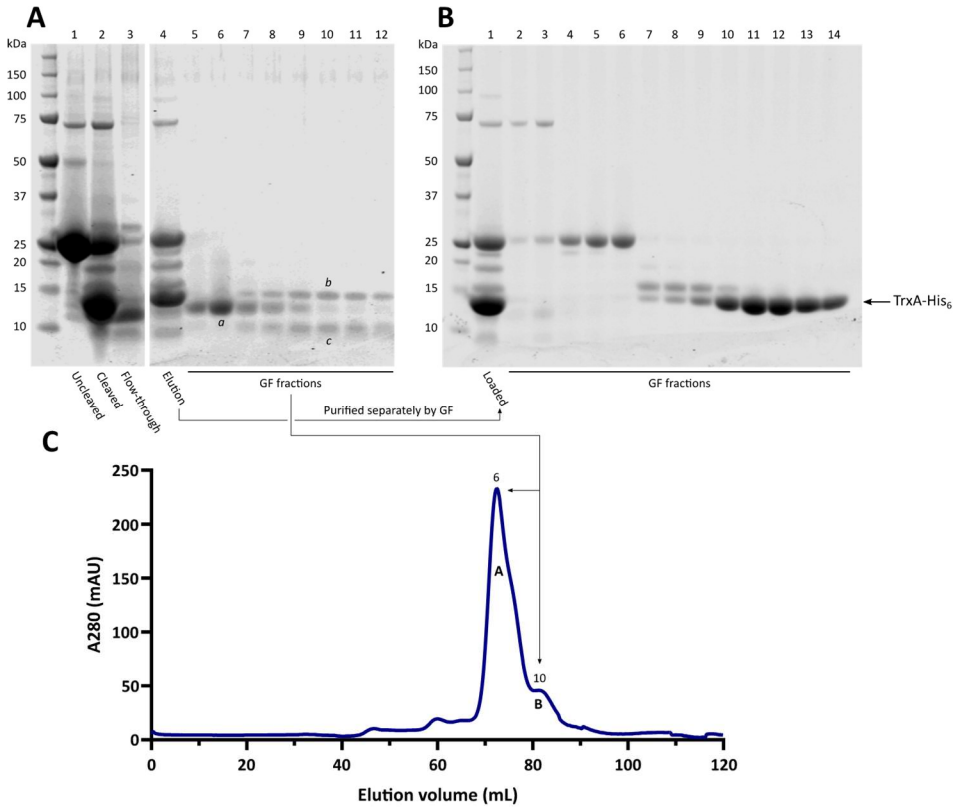


Figure 2.5: ^{15}N-labelled VWF C3 and TrxA-His_6 purification Panel A shows the analysis of fractions in the size exclusion chromatogram (Panel C). Lane 1 shows the uncleaved TrxA-His_6-VWFC3, Lane 2 the cleaved protein before Ni-NTA affinity chromatography, Lane 3 the unbound material which flowed through the Ni-NTA column and where VWF C3 would be present and Lane 4 the elution of the Ni-NTA where TrxA-His_6 would be present. Lanes 5-12 are the gel filtration fractions, with Lane 6 from the centre of Peak A and Lane 10 the centre of Peak B (Panel C). Bands *a*, *b* and *c* indicate bands that were analysed by mass spectrometry. Panel B analyses the size exclusion chromatography fractions from the purification of the elution fraction (Panel A, Lane 4). This was done to obtain a pure sample of TrxA-His_6. Lane 1 is the fraction loaded onto the gel filtration column with Lanes 2-14 the eluted gel filtration fractions. The 13 kDa band in Lanes 10-14 corresponds to TrxA-His_6.

The size exclusion chromatogram, following two-step Ni-NTA affinity chromatography and cleavage, presented two significant peaks, with Peak A eluting at approximately 72.5 mL and Peak B at 81.5 mL (Figure 2.5, Panel C). Peak A corresponded to a single band at 13 kDa (Figure 2.5, Panel A, Lanes 5-6), with Peak B presenting two bands at 15 kDa and 10 kDa (Lanes 10-12). Bands *a* (Figure 2.5, Panel A, Lane 6) (13 kDa), *b* and *c* (Lane 10) (15 and 10 kDa, respectively) were excised from the gel and sent to the St Andrews Mass Spectrometry Facility to be analysed by mass spectrometry. Each band was digested with trypsin and the tryptic peptides

of each gel band extracted. The digested peptides were analysed by liquid chromatography-mass spectrometry, an analytical technique which combines purification by liquid chromatography with analysis by mass spectrometry. This technique allowed the peptide sequences to be searched against a series of database protein sequences along with the 68-residue VWF C3 sequence. As expected, each band matched against both trypsin, used to digest the gel, and keratin, originating from skin, hair or dust particles which would have sat on the surface of the gel band, with high confidence. Two peptides purified from Band *a* (sequences VAQCSQKPCEDSCR and ETAAAKFER) and two peptides of Band *b* (sequences VAQCSQKPCEDSCR and SGFTYVLHEGECCGR) matched sequences found in the VWF C3 sequence. Each of these matches presented a high confidence score of 52 and 61 for the Band *a* sequences and 58 and 55 for the two Band *b* sequences. Only peptides matching the trypsin sequence displayed a higher confidence score with 69 and 81 in Band *a*, and 65, 74 and 76 in Band *b*. Peptides purified from Band *c*, however, did not match with any database peptide sequences or the VWF C3 sequence, only with the trypsin and once again with a high confidence score of between 52 and 61.

With both Bands *a* and *b* containing peptides which matched with the C3 sequence, a confident conclusion could be drawn that both gel bands contained the VWF C3 protein. It remained unclear why both bands were presenting a size far greater than 7.5 kDa. There is a possibility that the theoretical pI of the C3 domain being relatively acidic at 4.89 affects the electrophoretic mobility, with acidic proteins often displaying a greater observed size than its theoretical counterpart[405-407]. Another possibility is that a peptide impurity remains bound, likely covalently due to inaccurate cleavage considering the size observed is in reducing conditions. If inaccurate cleavage results in a small unstructured section of protein, the protein may appear larger when analysed by SDS-PAGE. Where this cleavage may take place remained unclear at this stage with no sites identified in the TrxA tag or linker connecting that to VWF C3 which are similar to the Ile-Glu-Gly-Arg FXa cleavage site. The 10 kDa impurity, Band *c*, bound to VWF C3 of Band *b* in solution could present further difficulty to the reliability of the structural assignment by altering the structure through its interactions and resulting in an inaccurate final VWF C3 structure. Whilst not ideal, if the protein size of Band *a* had resulted from inaccurate cleavage leaving extra amino acids in an unfolded 'tail' structure, this is less likely to interact with VWF C3 and alter its structure. These unstructured amino acids may, however, cause an issue with the protein assignment by overlapping with VWF C3 peaks, though this would not become clear until a later date. Furthermore, the protein concentration in that of Peak B may be sufficient for a ^{15}N-HSQC, though would be insufficient for the triple

resonance experiments needed for a structural assignment, with these pulse sequences far less sensitive than a ^{15}N-HSQC. Therefore, a ^{15}N-HSQC was carried out on the protein eluted in Peak A and this sample will henceforth be referred to as C3-^{15}N.

In order to confirm the cleavage from TrxA-His$_6$ was as expected, the elution fraction (Figure 2.5, Panel A, Lane 4) was separated using size exclusion chromatography. The SDS-PAGE of the eluted fractions presented a strong band at 13 kDa for TrxA-His$_6$ (Figure 2.5**,** Panel B, Lanes 11-14), smaller than the expected 17.3 kDa but consistent with the hypothesis that inaccurate cleavage by FXa had taken place. A ^{15}N-HSQC was carried out on this TrxA-His$_6$ sample, hereby referred to as TrxA-^{15}N. Alongside this, a ^{15}N-labelled sample of TrxA-His$_6$-VWFC3 was expressed, left uncleaved and a ^{15}N-HSQC performed. An individual peak in each ^{15}N-HSQC corresponds to an amide group, whether originating from the protein backbone or amino acid side chains. Initial observations that the peaks of each spectrum are well dispersed indicates that the protein is folded. Although it is unclear at this stage as to which peak corresponds to which amide group, such data is not required at this stage. Analysis of the spectra regarding overlap of peaks determines whether both C3-^{15}N and TrxA-^{15}N are the two cleaved fragments from the TrxA-His$_6$-VWFC3 protein.

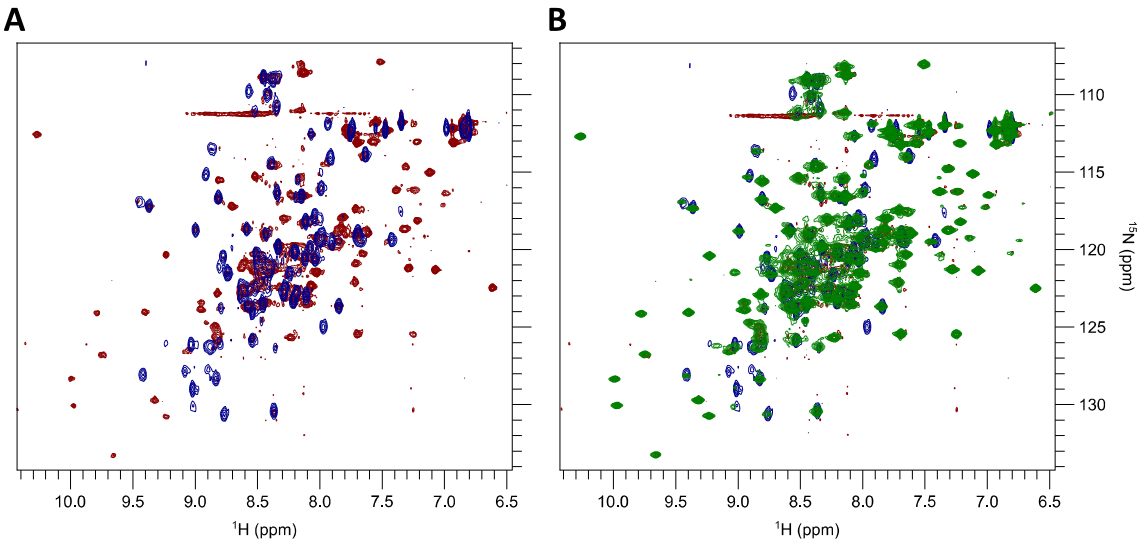


Figure 2.6: ^{15}N-HSQC overlay of VWF C3, TrxA-His$_6$ and the uncleaved TrxA-His$_6$-VWFC3 Panel A shows the ^{15}N-HSQC of both C3-^{15}N (blue) and TrxA-^{15}N (red). Overlaid with both spectra in Panel B is a ^{15}N-HSQC of uncleaved TrxA-His$_6$-VWFC3 (green). The uncleaved sample overlays well with both spectra with solely minor differences.

The overlay of the C3-^{15}N and TrxA-^{15}N with the uncleaved TrxA-His$_6$-VWFC3 ^{15}N-HSQC (Figure 2.6) confirmed that VWF C3 was present in the C3-^{15}N sample, with over 90% of well dispersed TrxA-^{15}N and C3-^{15}N peaks overlaying with that of the uncleaved TrxA-His$_6$-VWFC3. Those which failed to overlay likely stem from protein degradation, as seen around the ^{1}H chemical

shift of 8 ppm and ^{15}N chemical shift of 125 ppm. This is the region of the ^{15}N-HSQC where peaks from degraded protein tend to be observed. A second reason would be the new C-terminus of TrxA-^{15}N and N-terminus of C3-^{15}N following cleavage, which would cause peaks to shift due to their change in interactions with neighbouring residues. Whilst the question remained as to why the protein had a size of 13 kDa in reducing SDS-PAGE, the NMR results suggested that the purified protein was that of VWF C3.

To determine the optimal buffer conditions, the pH of the C3-^{15}N sample was changed from pH 7.4 to 6.9 and 6.4 and a ^{15}N-HSQC carried out in each condition. A buffer of 20 mM sodium phosphate, 100 mM NaCl, pH 7.4 was determined as the optimal conditions for protein assignment, with experiments carried out at 298 K. With triple resonance experiments used for a structural assignment often taking multiple days, especially considering the low concentration of purified VWF C3 (150-300 µM) compared to an ideal protein concentration for NMR assignment experiments (~1 mM), it was essential for the sample to remain stable with minimal degradation throughout these experiments. Any degradation is presented as small, sharp peaks in the 7.5-8 ppm/124-128 ppm (^{1}H/^{15}N) region. This is the region where C-terminal residues are observed, however, this would not be the case for VWF C3 as a proline residue is positioned at the C-terminus. The sample stability was tested with a ^{15}N-HSQC carried out every three days over a two-week period, with the sample at room temperature throughout the time course. Minimising degradation was an important facet throughout the expression process as the assignment experiments would be performed at room temperature. Any severe degradation would lead to a loss of signal and potential overlapping of peaks in the triple resonance experiments, however, the VWF C3 sample remained relatively stable with minimal degradation (Figure 2.7).

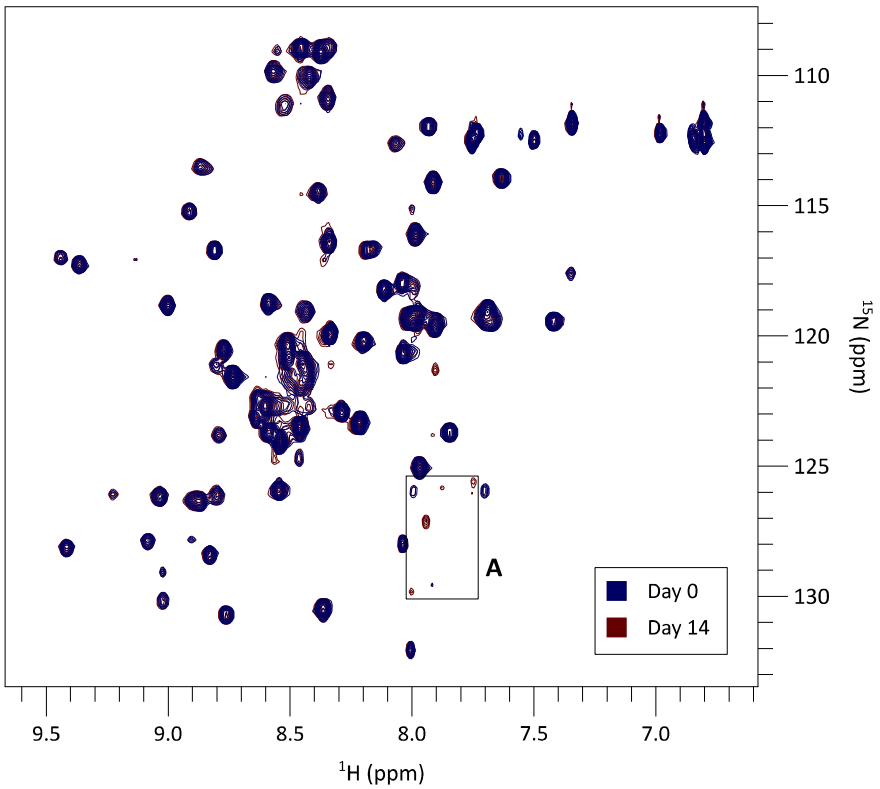


Figure 2.7: Stability of VWF C3 over a 2-week time course An overlay of the ^{15}N-HSQC spectra of C3-^{15}N on Day 0 (blue) and Day 14 (red). Minor degradation after 14 days can be observed by the small sharp red peaks in region A, however, minimal peak shift is observed throughout the spectrum, showing good protein stability.

2.3. Isotopically-labelled VWF C3 expression and purification

By confirming the presence of VWF C3 by mass spectrometry and ^{15}N-HSQC, a uniformly isotopically-labelled VWF C3 sample was expressed in M9 minimal media enriched with both $^{15}NH_4Cl$ and $^{13}C_6$-D-glucose. The sample was once again purified in a two-step Ni-NTA affinity chromatography protocol and subsequently applied to a size exclusion chromatography column (Figure 2.8).

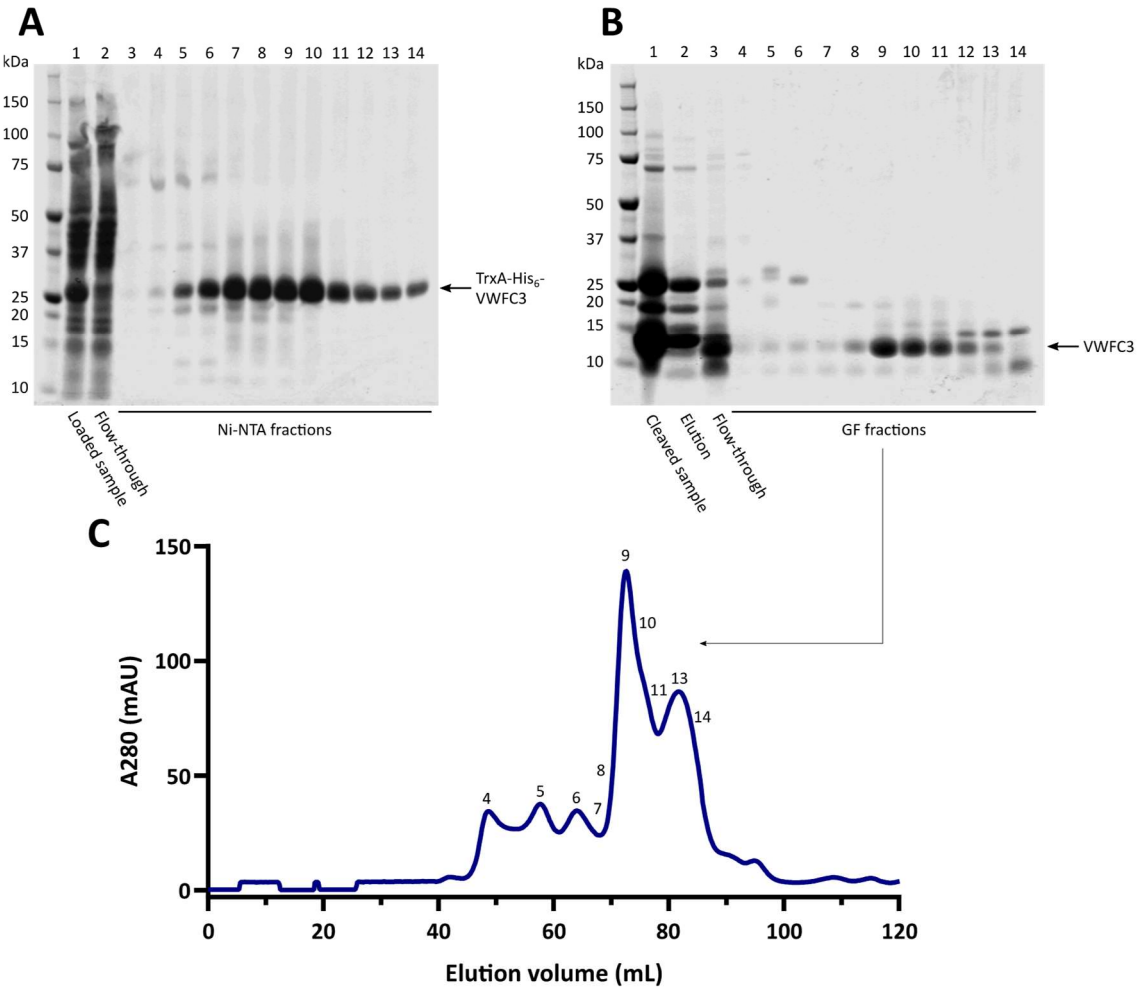


Figure 2.8: Isotopically-labelled VWF C3 purification Analysis of the uncleaved TrxA-His$_6$-VWFC3 Ni-NTA purification is shown in Panel A. Lane 1 shows the protein extracted from the *E. coli* cell cytoplasm with Lane 2 showing the unbound protein that flowed through the column. Lanes 3-14 present the eluted fractions with the strong band at 27 kDa in Lanes 5-14 corresponding to TrxA-His$_6$-VWFC3. Panel B shows the analysis of protein fractions eluted in the size exclusion chromatogram, following cleavage and a reverse Ni-NTA affinity column. Lane 1 shows the cleaved sample that was loaded onto the Ni-NTA column, Lane 2 displays the protein that bound to the column and was subsequently eluted, and Lane 3 shows the protein that flowed through the column and was loaded onto the size exclusion column. Lanes 4-14 analyse the eluted fractions under reducing SDS-PAGE, with each lane corresponding to the number indicated on the size exclusion chromatogram (Panel C).

The trace observed on the size exclusion chromatogram appeared to be reproducible with VWF C3 being eluted in one primary peak centred at 72.5 mL, corresponding to a single band of 13 kDa when analysed by SDS-PAGE. Alongside this, a second peak is again observed centred at 81.5 mL and migrates as two bands of 15 kDa and 10 kDa under reducing SDS-PAGE. Despite FXa cleavage once again not occurring in the intended position, the cleavage and purification traces did appear to be consistent and reproducible.

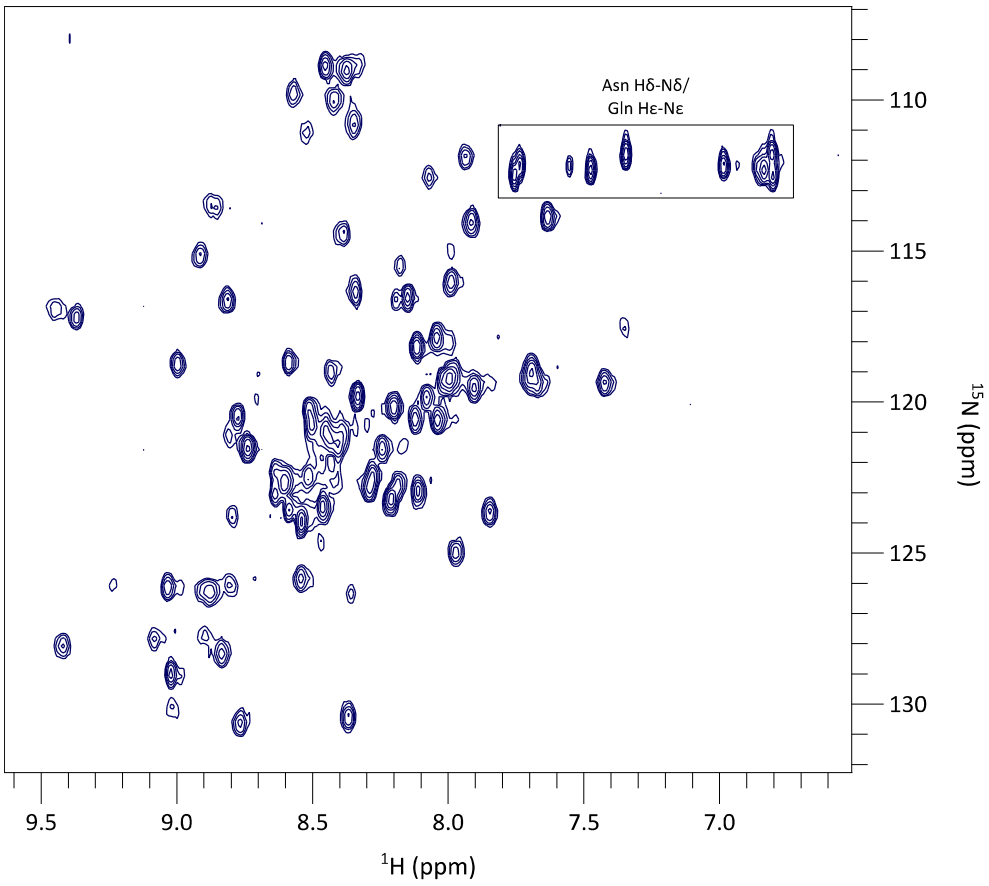


Figure 2.9: ^{15}N-HSQC of the isotopically-labelled VWF C3 A ^{15}N-HSQC of ^{15}N/^{13}C-labelled VWF C3. Indicated are peaks corresponding to the $^{15}NH_2$ side chain of asparagine ($^1H_\delta$-$^{15}N_\delta$) and glutamine ($^1H_\varepsilon$-$^{15}N_\varepsilon$). The spectrum was recorded on the 700 MHz spectrometer at 298 K in 20 mM sodium phosphate, 100 mM NaCl, pH 7.4.

The ^{15}N-HSQC spectrum presented 76 cross-peaks (Figure 2.9), when discounting those peaks corresponding to the asparagine and glutamine side chains. Furthermore, there would potentially be further peaks which are overlapped, particularly in the central region around the ^{1}H chemical shift of 8.5 ppm, or peaks which are missing due flexibility within the domain, leading to a high intrinsic rate of ^1H_N exchange. With three proline residues in VWF C3, this would equate to 65 cross-peaks, therefore, far more peaks were present than would be expected of a pure VWF C3 protein. This was, however, expected with the purified protein

presenting a size of 13 kDa in reducing SDS-PAGE, and inaccurate cleavage the most likely reason.

Correct formation of the five disulphide bonds in VWF C3 was also essential for accurate folding of the domain. An Ellman's Assay was able to confirm that all cysteines were involved in disulphide bonds with a free thiol concentration of -1.1 µM (Table 2.1) in the VWF C3 sample, indicating that no free thiols were present in the purified protein.

Standard	Free thiol concentration (mM)	OD_{412}
A	1.5	1.898
B	1.25	1.591
C	1	1.279
D	0.75	0.957
E	0.5	0.644
G	0.25	0.318
F	0	0
VWF C3	**-0.0011**	**0.003**

Table 2.1: Ellman's Assay for the free thiol concentration of the expressed protein Displayed is the Ellman's Assay result, showing the free thiol concentration and OD_{412} values of each standard (A-F). Displayed in bold is the free thiol concentration of the expressed C3 protein, calculated using the trendline equation of the standards, $y = 1.2679x + 0.0044$, which had an R^2 value of 0.9999.

A folded protein could be confidently assumed by the ^{15}N-HSQC, with peaks in the ^{1}H dimension well dispersed throughout the spectrum. Unfolded and unstructured proteins display poorly dispersed peaks with the $^{1}H_N$ chemical shift concentrated in the 7.5-8.5 ppm range, due to their lack of interaction with neighbouring residues and all residing in a similar chemical environment. The importance of disulphide bond formation in forming folded VWF C3 was displayed by the significant change in the ^{15}N-HSQC upon reduction of the domain by DTT (Figure 2.10), causing peaks to be concentrated in the 7.5-8.5 ppm range versus the 6.5-9.5 ppm range in folded VWF C3.

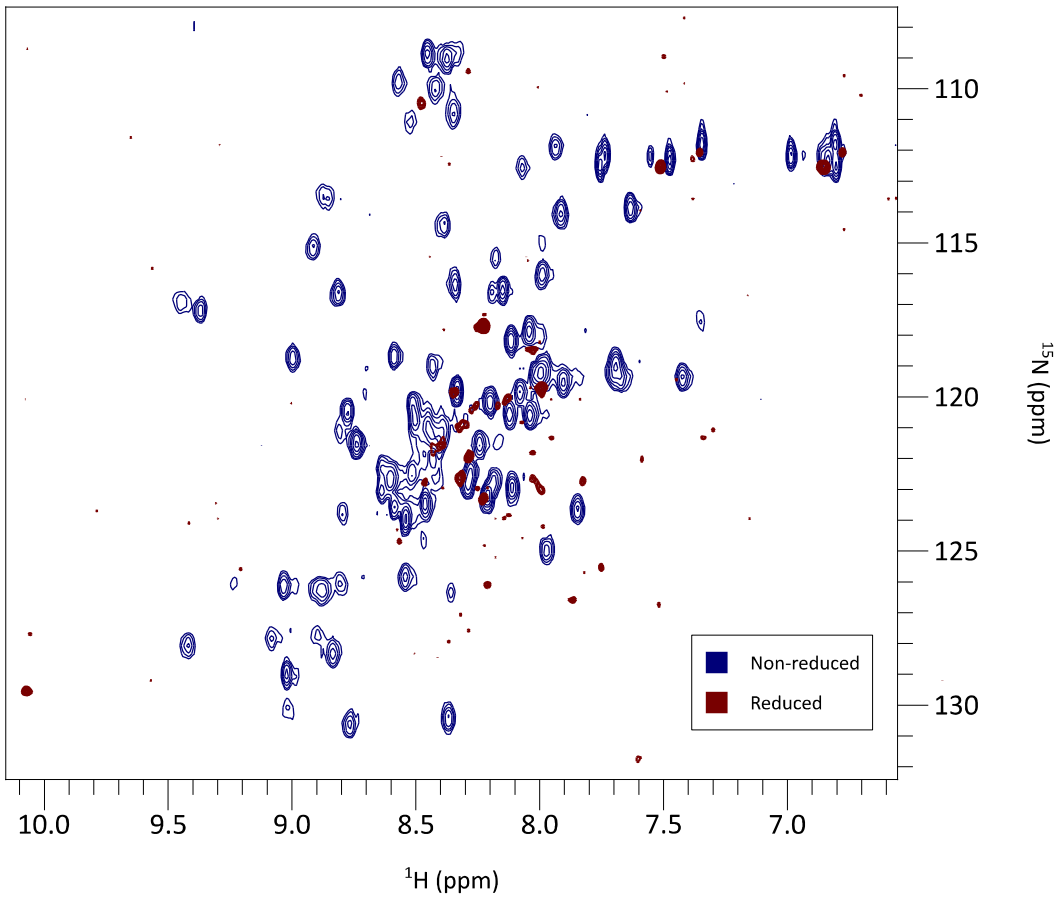


Figure 2.10: ^{15}N-HSQC of reduced VWF C3 A ^{15}N-HSQC of reduced VWF C3 (red) compared to that of non-reduced VWF C3 (blue). Both experiments were performed on a 700 MHz spectrometer in 20 mM sodium phosphate, 100 mM NaCl, pH 7.4, with the reduced sample also containing 10 mM DTT.

With conditions optimised and an isotopically-labelled C3 sample expressed, triple resonance experiments were carried out to assign the VWF C3 amide backbone.

2.4. NMR triple resonance experiments and backbone amide assignment

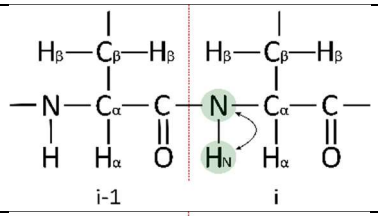
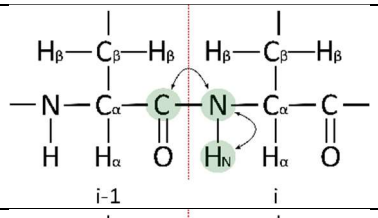
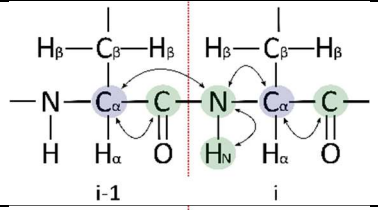
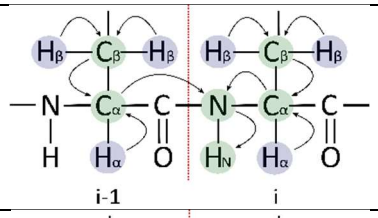
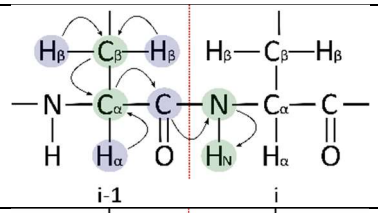
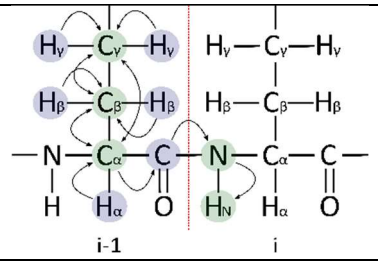
Experiment	Magnetisation transfer	Correlation
^{15}N-HSQC (2D)		$^{1}H_N^{i}$-$^{15}N_H^{i}$
HNCO (3D)		$^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_O^{i-1}$
HN(CA)CO (3D)		$^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_O^{i-1}$ $^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_O^{i}$
HNCACB (3D)		$^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_\alpha^{i-1}$ $^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_\beta^{i-1}$ $^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_\alpha^{i}$ $^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_\beta^{i}$
HN(CO)CACB (3D)		$^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_\alpha^{i-1}$ $^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_\beta^{i-1}$
CC(CO)NH-TOCSY (3D)		$^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_\alpha^{i-1}$ $^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_\beta^{i-1}$ $^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_\gamma^{i-1}$ $^{1}H_N^{i}$-$^{15}N_H^{i}$-$^{13}C_\delta^{i-1}$ Correlation shown for all ^{13}CH groups in the side chain

Table 2.2: Experiments used for the backbone $^{1}H_N^{15}N_H$ assignment The experiments used in the assignment of the backbone amide group, their magnetisation transfer and the resulting correlations observed. Letters in the experiment names correspond to the nuclei whose chemical shift is correlated, with those in brackets involved in the coherence transfer pathway. Nuclei highlighted in blue indicate where magnetisation is excited/transferred and green where the chemical shift is evolved. An individual residue is denoted by *i* with its preceding residue labelled *i-1*.

The experiments outlined (Table 2.2) were used for the backbone $^{1}H_N^{15}N_H$ assignment, with the HNCACB and HN(CO)CACB being the spectra used primarily. Whilst not essential for an assignment at this stage, the CC(CO)NH-TOCSY experiment was a useful addition to these,

displaying cross-peaks for each aliphatic ^{13}CH, $^{13}CH_2$ and $^{13}CH_3$ group in the side chain which can provide information on the preceding residue type. This can be particularly useful for residues with a unique side chain ^{13}C chemical shift pattern such as arginine or isoleucine, though a cross-peak is not necessarily displayed for every aliphatic ^{13}C due to weaker signal further along the side chain. The HNCO and HN(CA)CO experiments were used when the signal was particularly low in the other triple resonance experiments, with the HNCO being the most sensitive of all bar the ^{15}N-HSQC.

Each peak in the ^{15}N-HSQC was assigned a number known as a spin system, with each spin system corresponding to one amino acid. The spin system of each peak was also assigned to the cross-peaks in the corresponding $^{1}H_N{}^{15}N_H$ strip of each triple resonance experiment, designating both $^{13}C_O{}^{i}$, $^{13}C_O{}^{i-1}$, $^{13}C_\alpha{}^{i}$, $^{13}C_\alpha{}^{i-1}$, $^{13}C_\beta{}^{i}$ and $^{13}C_\beta{}^{i-1}$ (Figure 2.11). This was straightforward for those peaks that were well dispersed, however, more challenging for those heavily overlapped peaks in the central region of the ^{15}N-HSQC.

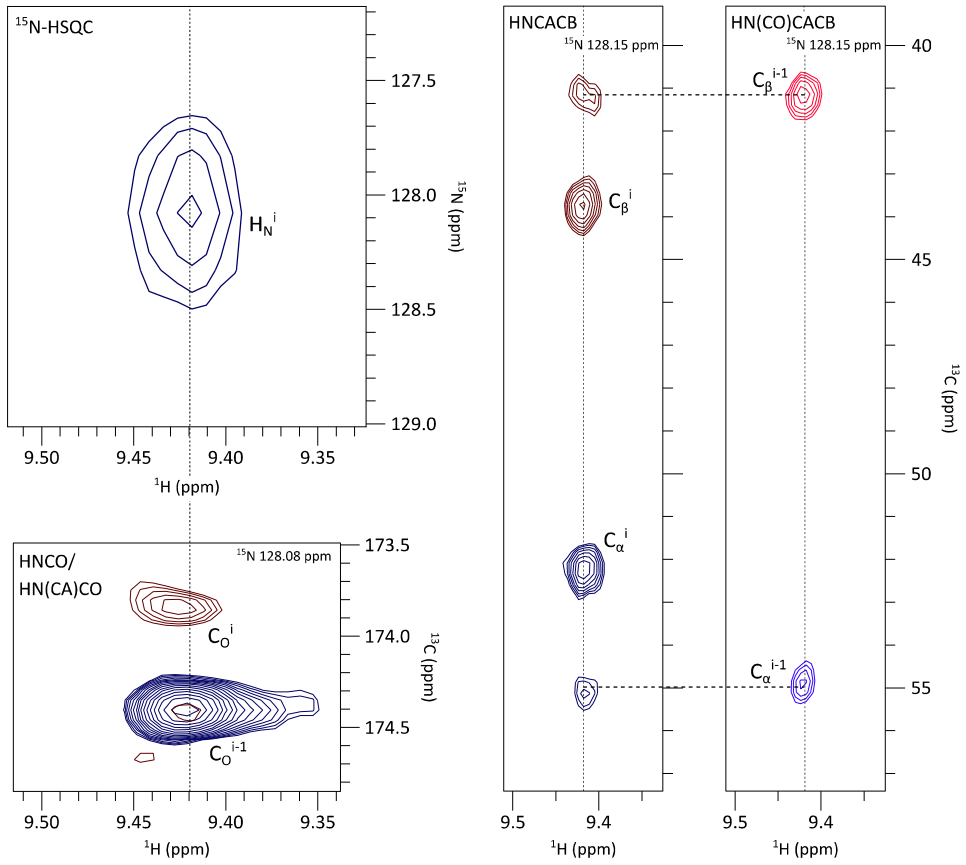


Figure 2.11: Spin system strips in the backbone assignment spectra A comparison of the correlated strips observed in five NMR spectra for the assigned L2495 residue (residue *i*). The peak observed in the ^{15}N-HSQC is correlated to that in the HNCO (blue) and HN(CA)CO (red) spectra in both the ^{1}H (9.419 ppm) and ^{15}N (128.08 ppm) dimensions, which subsequently provide the $^{13}C_O$ chemical shift of L2495 (*i*) and C2494 (*i-1*). These also correlate to the ^{1}H and ^{15}N dimensions of the HNCACB and HN(CO)CACB to display the $^{13}C_\alpha$ (blue) and $^{13}C_\beta$ (red) chemical shifts of L2495 and C2494.

With both HNCACB and HN(CO)CACB providing information on $^{13}C_{\alpha}^{i}$, $^{13}C_{\alpha}^{i-1}$, $^{13}C_{\beta}^{i}$ and $^{13}C_{\beta}^{i-1}$, a sequential assignment was possible primarily with these spectra (Figure 2.12). For each $^{1}H_{N}^{i}$ strip in the HNCACB spectrum, 4 peaks are observed with $^{13}C_{\alpha}$ peaks positive and $^{13}C_{\beta}$ peaks negative, or vice versa depending on the phasing. The exception to this is when glycine is either residue *i* or *i-1* as it only contains a single $^{13}C_{\alpha}$ peak. Overlaying with the HN(CO)CACB spectrum discloses which $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ pair are from *i* and *i-1*, with the $^{13}C_{\alpha}^{i}$ and $^{13}C_{\beta}^{i}$ only observed in the HNCACB. It is usually the case that the $^{13}C_{\alpha}^{i}$ and $^{13}C_{\beta}^{i}$ peaks show a stronger signal though this is not always the case. There may also be a possibility that the chemical shifts of $^{13}C_{\alpha}^{i}$ and $^{13}C_{\alpha}^{i-1}$, or $^{13}C_{\beta}^{i}$ and $^{13}C_{\beta}^{i-1}$ are very similar, leaving them overlapped and only one peak observed. This is particularly likely where identical residues are seen in sequence, such as C2490 and C2491 in VWF C3.

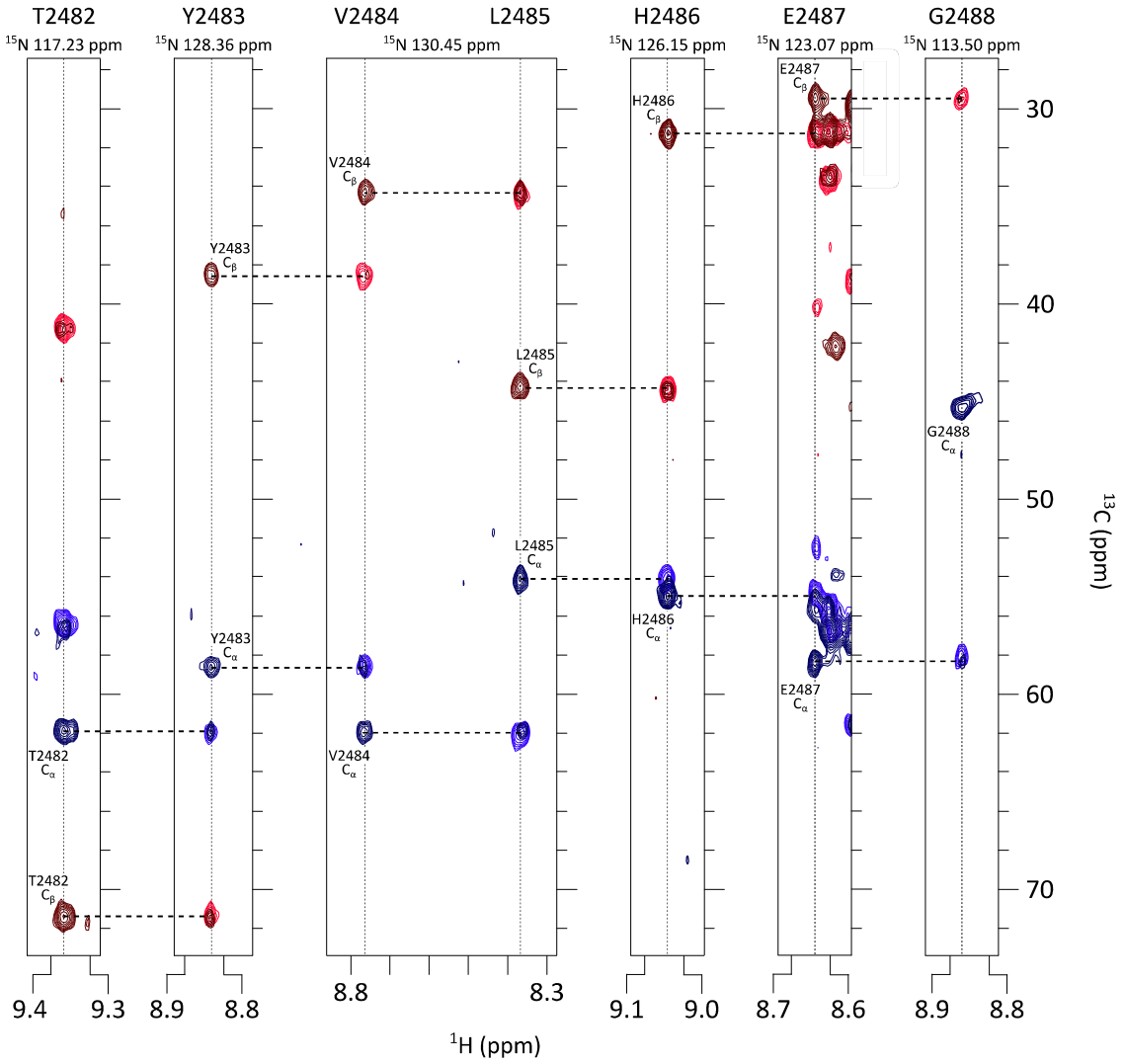


Figure 2.12: Sequential assignment using the HNCACB and HN(CO)CACB spectra The assignment of the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts of T2482 to G2488, and subsequently the backbone amide chemical shifts, using the HNCACB ($^{13}C_{\alpha}$ – dark blue, $^{13}C_{\beta}$ – dark red) and HN(CO)CACB ($^{13}C_{\alpha}$ – light blue, $^{13}C_{\beta}$ – light red) experiments. Connections between the $^{13}C_{\alpha}^{i-1}$/$^{13}C_{\beta}^{i-1}$ to the $^{13}C_{\alpha}$/$^{13}C_{\beta}$ of the preceding residue's strip are indicated.

The assignment by this method provided a number of sequential assignments of spin systems and led to the assignment of a small number of individual residues, such as alanine which has a uniquely low $^{13}C_{\beta}$ chemical shift with an average of 18.96 ppm, or serine and threonine which have a larger $^{13}C_{\beta}$ chemical shift than $^{13}C_{\alpha}$. However, few assignments of individual residues were possible due to the difficulty in differentiating between residues using only $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts. This is where the CC(CO)NH-TOCSY spectrum proved particularly useful, with certain residues displaying unique ^{13}C side chain chemical shift patterns, such as isoleucine whose $^{13}C_{\gamma 1}$, $^{13}C_{\gamma 2}$ and $^{13}C_{\delta 1}$ all display a chemical shift on average between 28 and 13 ppm. It also differentiated between serine and threonine residues which display a similar $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shift pattern but threonine displays a $^{13}C_{\gamma 2}$ cross-peak at approximately 21 ppm. This, however, required the signal to be strong enough for these cross-peaks to be observed as the CC(CO)NH spectrum is less sensitive than the HNCACB experiment.

In cases where the HNCACB/HN(CO)CACB signal was too weak for a sequential assignment between *i* and *i-1*, the HN(CA)CO/HNCO spectra were used to determine the $^{1}H_{N}$ and $^{15}N_{H}$ chemical shift (Figure 2.13). In each $^{1}H_{N}^{i}$ strip, the $^{13}C_{O}^{i}$ and $^{13}C_{O}^{i-1}$ cross-peaks are observed in the HN(CA)CO experiment, with the $^{13}C_{O}^{i}$ generally the stronger of the two, and only the $^{13}C_{O}^{i-1}$ cross-peak observed in the HNCO. All backbone $^{13}C_{O}$ cross-peaks were present in the range of 171-178.5 ppm.

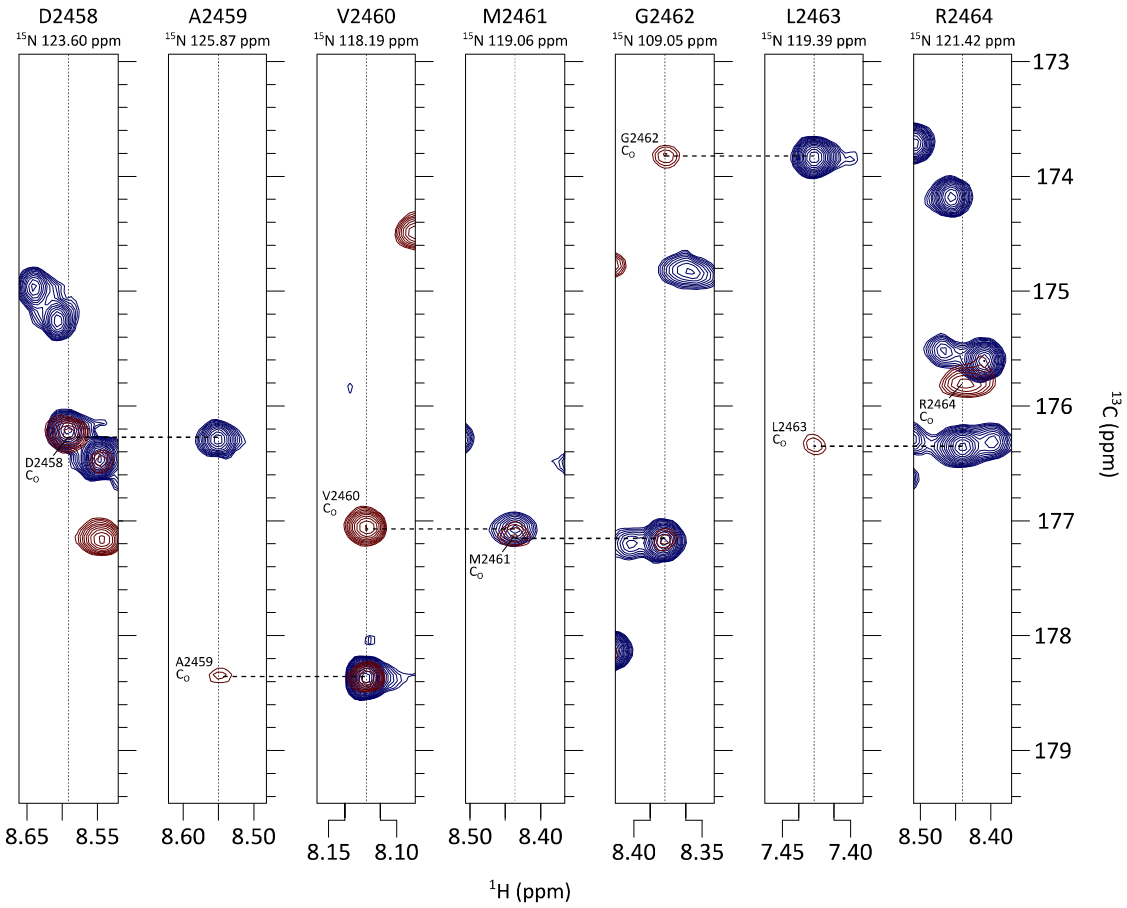
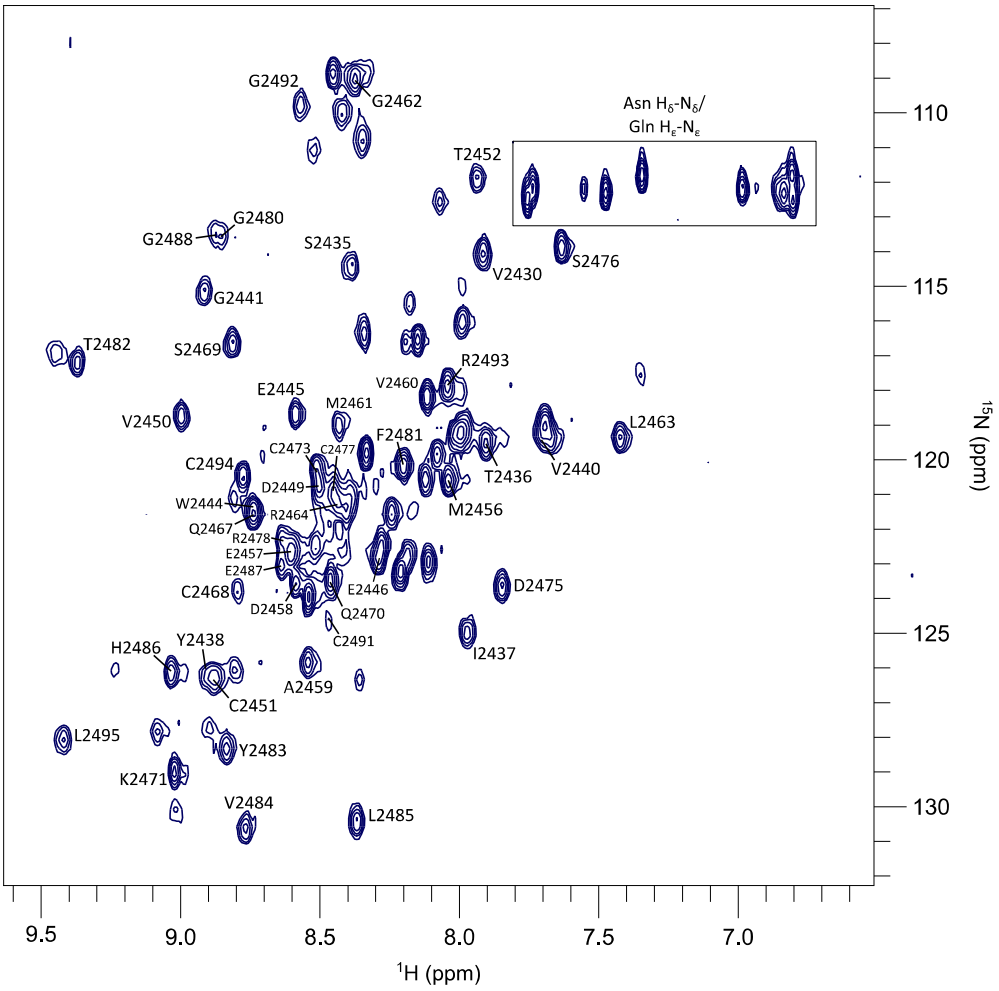


Figure 2.13: Sequential assignment using the HNCO/HNCACO experiments The assignment of the $^{13}C_O$ chemical shift of D2458 to R2464 using the HNCO (blue) and HN(CA)CO (red) experiments. Connections between $^{13}C_O^{i-1}$ peaks to the CO of the preceding residue's strip is indicated.

2.5. Partial backbone assignment and linker impurity

Though a backbone $^{1}H_{N}{}^{15}N_{H}$ assignment was completed for 47 of 68 residues, issues arose throughout the process with a lack of signal in the triple resonance spectra in certain regions of the protein. An assignment was completed for the residues indicated in Figure 2.14.



KVCVHR**STIY**P**VG**QF**WEE**GC**DVCT**CTD**MEDAVMGLR**VA**QCSQK**P**C**E**DSCRSGFTYVLHEG**EC**CGRCL**P

Figure 2.14: Partial backbone amide assignment of the VWF C3 domain The VWF C3 assignment of the ^{15}N-HSQC, with 47 of 68 amide residues assigned. Indicated below the spectrum in blue are those residues whose the backbone amide chemical shift was assigned.

Discounting three proline residues, 18 residues remained unassigned. Particularly worrying was the lack of assignment possible using the HN(CA)CO/HNCO spectra as these are the most sensitive triple resonance experiments. Lack of backbone $^{1}H_{N}{}^{15}N_{H}$ assignment would likely indicate further issues in the case of a side chain assignment using less sensitive experiments. Considering five disulphide bonds in VWF C3, it is unlikely that the lack of signal in these regions is due to them being highly flexible, particularly as a number of these regions are

centred around cysteine residues (C2431, C2448, C2453, C2473, C2490). This is more likely due to microsecond to millisecond dynamics which causes line broadening in cysteine residues[408].

A further issue that arose, whilst hypothesised in the purification of VWF C3, was the presence of a protein impurity originating from the linker connecting the His_6 tag to VWF C3, with assigned residues (red) indicated below.

SSGLVPRGSG**M**K**ETAAAKF**ERQHM**DS**P**DLGTGGGSGIEG**R

The presence of this impurity became apparent by the assignment of a glycine residue preceded by a threonine and a threonine preceded by a glycine. Both display unique and distinguishable ^{13}C chemical shift patterns with the $^{13}C_\beta$ chemical shift of threonine more positive than $^{13}C_\alpha$ and a cross-peak at approximately 21 ppm corresponding to the side chain $^{13}C_{\gamma 2}$. Combining this with a glycine presenting a single $^{13}C_\alpha$ cross-peak at 45 ppm, the identity of these residues was unmistakeable. Also apparent was the number of glycine residues in consecutive sequence in the ^{15}N region of 110 ppm, as is only found in the linker region. Though likely due to the single 13 kDa band in reducing SDS-PAGE (Figure 2.8, Panel B, Lane 9), it was not possible to confirm that the linker was bound due to incomplete FXa cleavage. If the hypothesis was correct, a sequential assignment would be possible from K2429 to the arginine residue at the C-terminus of the cleavage site, or vice versa, however this was not possible as an assignment of neither of these had been confirmed. In order to determine whether the linker impurity remain bound to VWF C3, a diffusion experiment was carried out in which a diffusion coefficient of each of the assigned peaks in the ^{15}N-HSQC was calculated (Figure 2.15).

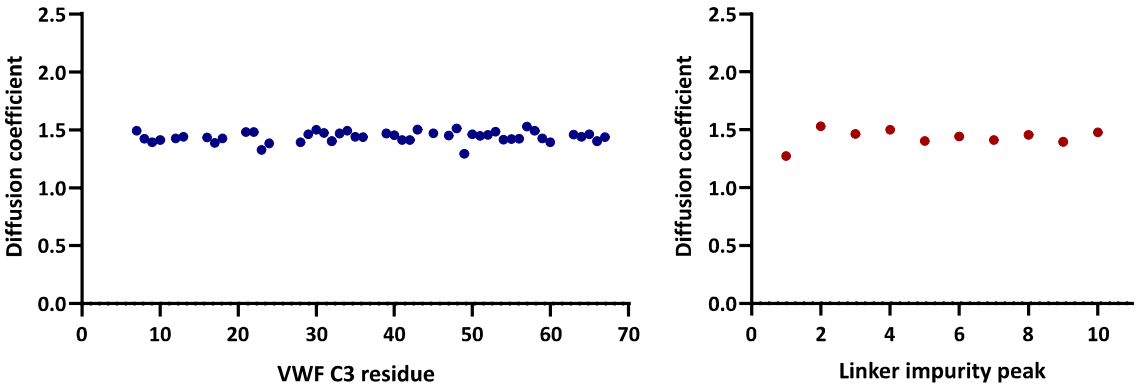


Figure 2.15: Diffusion coefficients of VWF C3 and the linker impurity A comparison between the diffusion coefficients of the 47 assigned VWF C3 residues (blue) and 10 assigned residues of the assigned peaks from the linker impurity (red). The similarity between diffusion coefficients of the VWF C3 residues and linker impurity, with all values in the 1.27 to 1.53 range, the two are likely bound in solution.

The diffusion experiment was performed using the DOSY experiment[409]. If the linker impurity was bound to VWF C3, all assigned peaks would display a similar diffusion coefficient, however, if both were unbound and diffusing independently, their diffusion coefficients would differ. A two-tailed t-test was used to compare the diffusion coefficients of the peaks from VWF C3 and the linker impurity. This ascertained whether the means of the two groups were statistically similar enough to be from the same population. The t-test indicated a probability of 0.832 that the two sets of data originated from the same population. This allowed a confident assumption that the two were covalently bound and further hypothesising that inaccurate cleavage had occurred. Why or where this inaccurate cleavage had occurred remained unclear though the explanation became clear at a later date (see Chapter 3.9).

2.6. VWF C3 (EK) plasmid design and expression

Due to the issue of the linker impurity and the low purified VWF C3 protein concentration, a new plasmid was designed for expression of the VWF C3 domain.

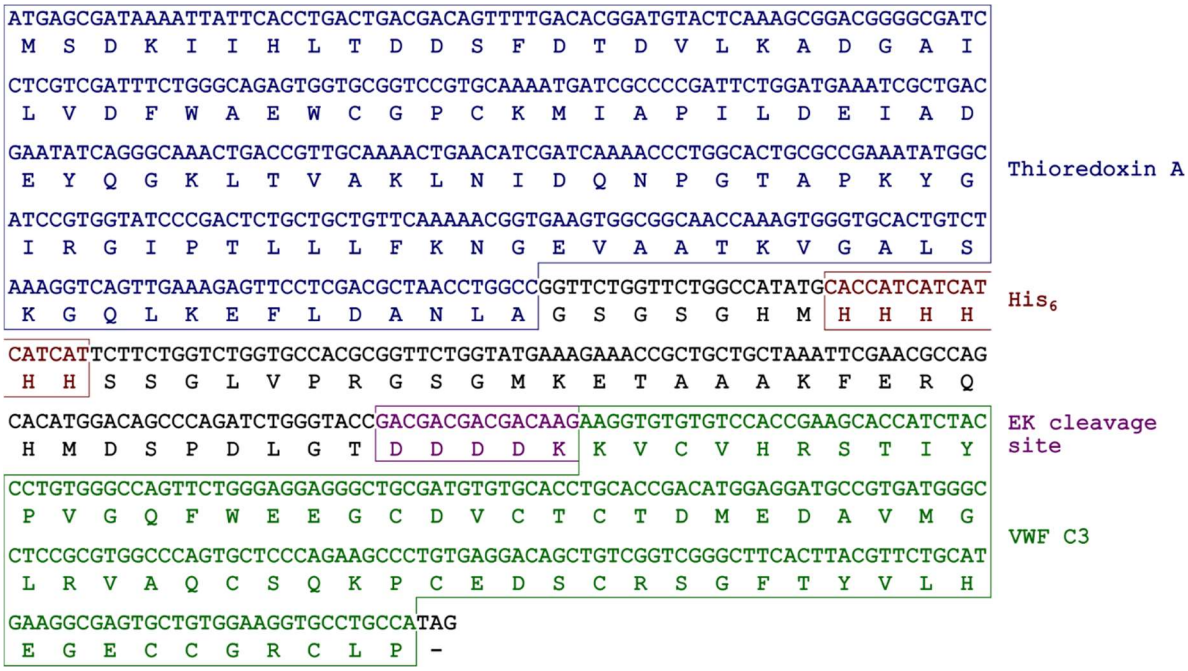


Figure 2.16: TrxA-His$_6$-VWFC3 (EK) construct sequence The nucleotide sequence and subsequent translation product of the newly designed plasmid, featuring an EK cleavage site. The 109 amino acid TrxA tag (blue) is once again separated by a His$_6$ tag (red) by 7 residues, with the His$_6$ tag followed by a 31 amino acid linker to an EK cleavage site (purple). The cleavage site is immediately followed by the VWF C3 protein sequence (green) which is subsequently followed by a termination codon.

An enterokinase (EK) cleavage site (Asp-Asp-Asp-Asp-Lys$^{\downarrow}$) was deemed most suitable instead of FXa. It was hypothesised that FXa was too non-specific to cleave in the correct position and as TEV was not suitable due to its reduced efficiency in oxidising conditions, EK was chosen as the protease. A pET-32b(+) plasmid was once again used as a template, with a 31-residue linker between the His_6 tag and the EK cleavage site, the same linker as the previous plasmid, however, minus the Gly-Gly-Gly-Ser-Gly at its C-terminus. The VWF C3 sequence was placed immediately following the EK cleavage site and followed at its C-terminus by a stop codon (Figure 2.16).

The protein was expressed and purified in the same manner as previously, expressed in M9 minimal media enriched with $^{15}NH_4Cl$ and unlabelled $^{12}C_6$-$^{1}H_{12}$-glucose and the uncleaved TrxA-His_6-VWFC3 purified using Ni-NTA affinity chromatography. One factor that became immediately apparent upon purification of the uncleaved protein was the vastly superior protein concentration, as when expressed in two litres of M9 minimal media, the UV trace peaked at 4500 mAU eluting over a 50 mL volume. This was compared to a peak of approximately 1000 mAU in the previous plasmid eluting over a similar volume, expressed in four litres of M9 minimal media (Figure 2.17, Panel A). Whilst this reading of 4500 mAU may be due to an anomaly caused by air trapped in the AKTA system, when analysed by reducing SDS-PAGE, it was clear that the newly designed plasmid expressed far superior protein concentration. The reason for this remains unclear, with the only difference between the two plasmids being the removal of five residues in the His_6-EK site linker and the change in the cleavage site itself. This may also be caused by a difference elsewhere in the plasmid which allows *E. coli* to express the protein more efficiently.

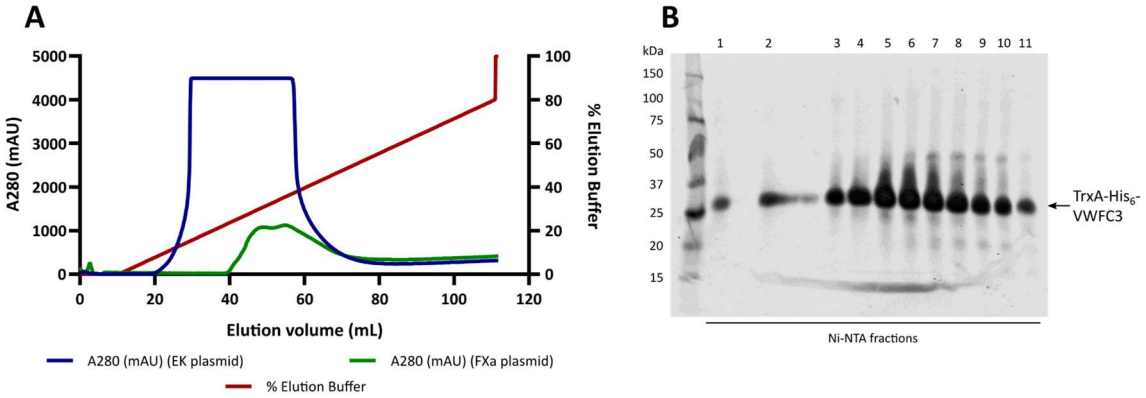


Figure 2.17: Purification of the uncleaved TrxA-His_6-VWFC3 (EK) Panel A shows the purification of the TrxA-His_6-VWFC3 (EK) by Ni-NTA (red) in comparison to that of TrxA-His_6-VWFC3 (FXa) (blue). Analysis of the TrxA-His_6-VWFC3 (EK) elution fractions in reducing SDS-PAGE is shown in Panel B. All lanes, 1-11, are fractions eluted through the Ni-NTA affinity chromatography column, with the large band at 27 kDa corresponding to TrxA-His_6-VWFC3.

2.7. Cleavage optimisation, purification and NMR analysis

A test was once again carried out on the optimal cleavage time, with cleavage undertaken at 18°C rather than 4°C. This was decided as EK cleaves more efficiently at room temperature, therefore, whilst minor protein degradation may have occurred, a more efficient cleavage would result in increased VWF C3 protein concentration. Buffer conditions were also changed, with the cleavage and purification carried out in 20 mM Tris, 50 mM NaCl, pH 7.4 as EK cleaves poorly in sodium phosphate buffer.

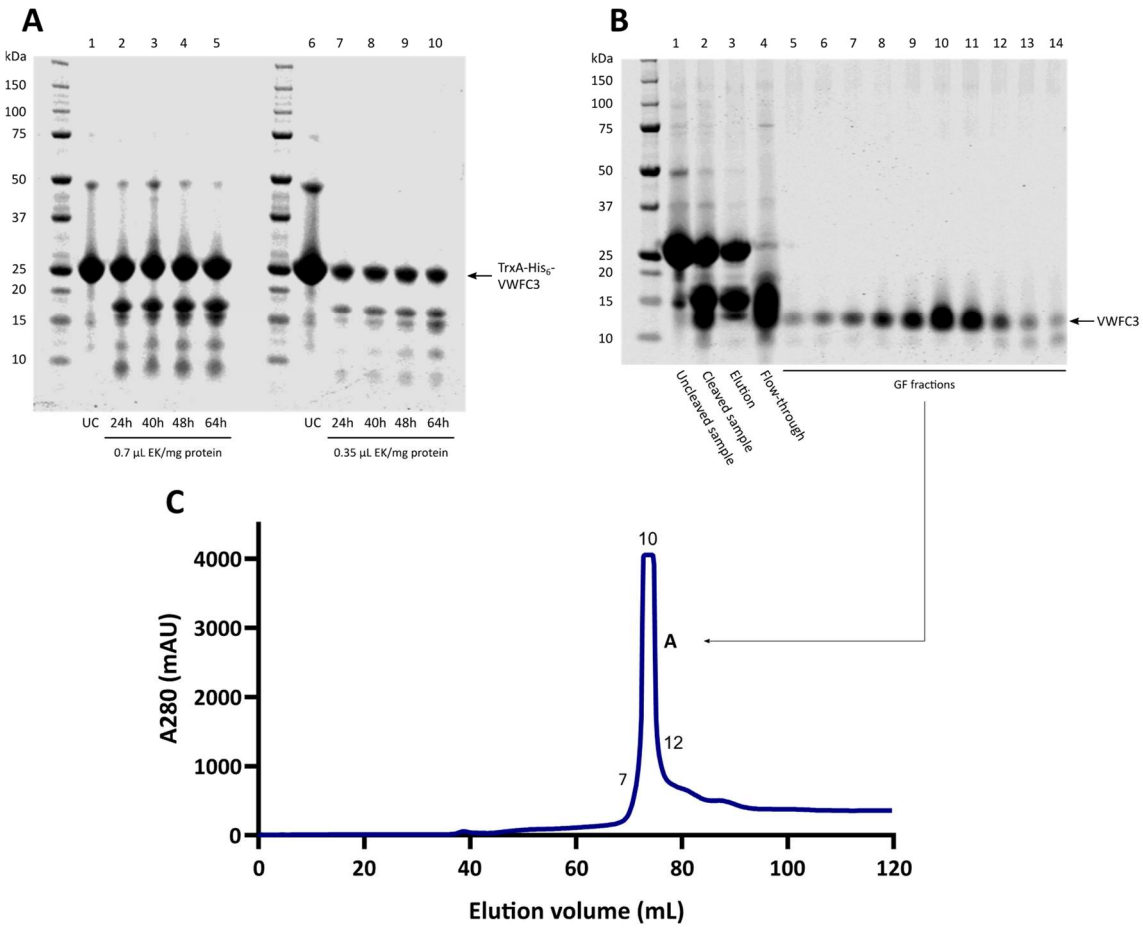


Figure 2.18: Small-scale cleavage and purification of VWF C3 A small-scale cleavage of ^{15}N-labelled TrxA-His$_6$-VWFC3 is displayed in Panel A, with Lanes 2-5 analysing cleavage from 24 to 64 hours at 0.7 µL EK/mg protein. Lanes 7-10 analyse cleavage with 0.35 µL EK/mg protein over the same time points. With 40 hours at 0.7 µL EK/mg protein chosen as most efficient, Panel B and C show purification of cleaved VWF C3 by size exclusion chromatography following a reverse Ni-NTA. The uncleaved TrxA-His$_6$-VWFC3 sample is shown in Lane 1 (Panel B), the cleaved sample that was loaded onto the Ni-NTA column in Lane 2 and the bound protein that was subsequently eluted by 500 mM Imidazole shown in Lane 3. The unbound protein that flowed through the column in displayed in Lane 4, with Lanes 5-14 corresponding to the fractions eluted in the size exclusion chromatogram (Panel C) and Lanes 7, 10 and 12 from the positions indicated.

Cleavage of TrxA-His_6-VWFC3 was determined as most efficient at 40 hours at a concentration of 0.7 µL EK/mg protein, all carried out shaking at 40 rpm at 18°C. Once cleaved, the protein was purified as before with a Ni-NTA affinity and size exclusion chromatography (Figure 2.18). When comparing between protein expressed from the previous plasmid and the new plasmid, that expressed from the previous plasmid will henceforth be referred to as VWF C3 (FXa), that expressed from the new plasmid will be VWF C3 (EK).

Whilst the eluted VWF C3 (EK) appeared to be in significantly higher concentrations and relatively pure, the same issue seemingly occurred as with VWF C3 (FXa), with the VWF C3 (EK) protein shown as a single band of 13 kDa in reducing SDS-PAGE. This again suggests that cleavage did not occur at the intended site, however, would not become clear until the completion of the backbone amide assignment. The purified VWF C3 (EK) protein concentration was far superior, with the UV absorbance peaking at 4500 mAU, compared to 230 mAU for the ^{15}N-labelled protein of VWF C3 (FXa). Despite cleavage likely not occurring at the intended site, the improved protein concentration would likely make a full backbone assignment possible, with the only potential issue being overlap of peaks from the His_6-VWFC3 linker if, as suspected, cleavage had occurred somewhere within the linker. The pure sample from Peak A was collected, analysed by a ^{15}N-HSQC and compared to the ^{15}N-HSQC of VWF C3 (FXa) (Figure 2.19).

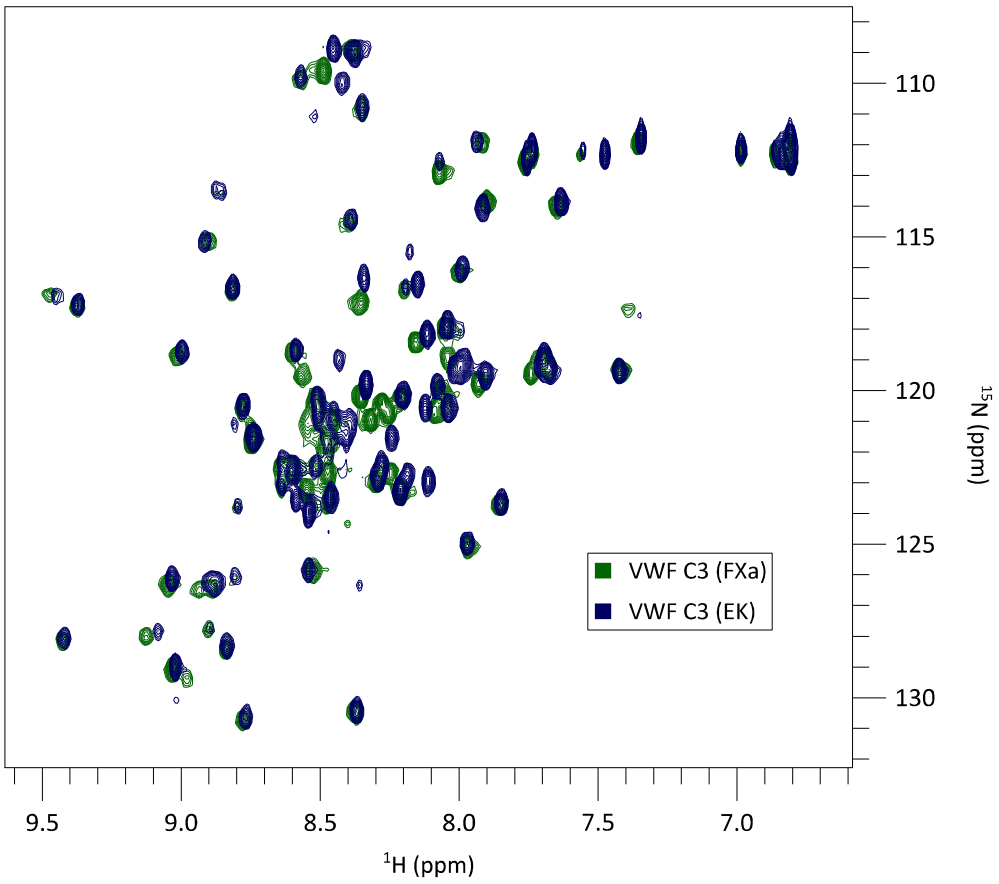


Figure 2.19: Comparison of VWF C3 (EK) ^{15}N-HSQC to VWF C3 (FXa) A comparison of the VWF C3 ^{15}N-HSQC spectrum purified by the original (blue) and new (green) plasmid. Both spectra were recorded in 20 mM sodium phosphate, 100 mM NaCl, pH 7.4 at 298 K, however, VWF C3 (EK) was recorded on the 600 MHz spectrometer and VWF C3 (FXa) was recorded on the 700 MHz spectrometer. VWF C3 (EK) (green) overlays well with the VWF C3 (FXa) (blue) spectrum, with some peak shift as would be expected with a new plasmid and different spectrometer.

The peaks of VWF C3 (EK) matched well to the assigned peaks of VWF C3 (FXa) with some minor peak shift but as would be expected in a new sample. There were 82 peaks displayed for the VWF C3 (EK) protein, once again far more than what would be expected for a pure VWF C3 sample, increasing the likelihood that inaccurate cleavage had occurred. There appeared to be fewer peaks presented in the 8.5 ppm (^{1}H)/110 ppm (^{15}N) region, likely due to the loss of the Gly-Gly-Gly-Ser-Gly residues that were present in the linker of the VWF C3 (FXa) plasmid. The ^{15}N-HSQC of the VWF C3 (EK) sample was in the same sodium phosphate buffer as VWF C3 (FXa), however, was recorded on the 600 MHz spectrometer, rather than the 700 MHz which VWF C3 (FXa) was recorded which would also contribute to the minor shifting of peaks. Optimal buffer conditions were tested, with ^{15}N-HSQC carried out on a range of conditions including 20 mM Tris vs 20 mM sodium phosphate buffer, 50 mM vs 100 mM NaCl, pH 7.4 vs 6.4. It was decided that the optimal conditions were 20 mM sodium phosphate, 50 mM NaCl, pH 6.4, with the sample showing good stability at room temperature with minimal peak shift

over a two-week time course (Figure 2.20). All assignment experiments henceforth were carried out in these buffer conditions at 298 K.

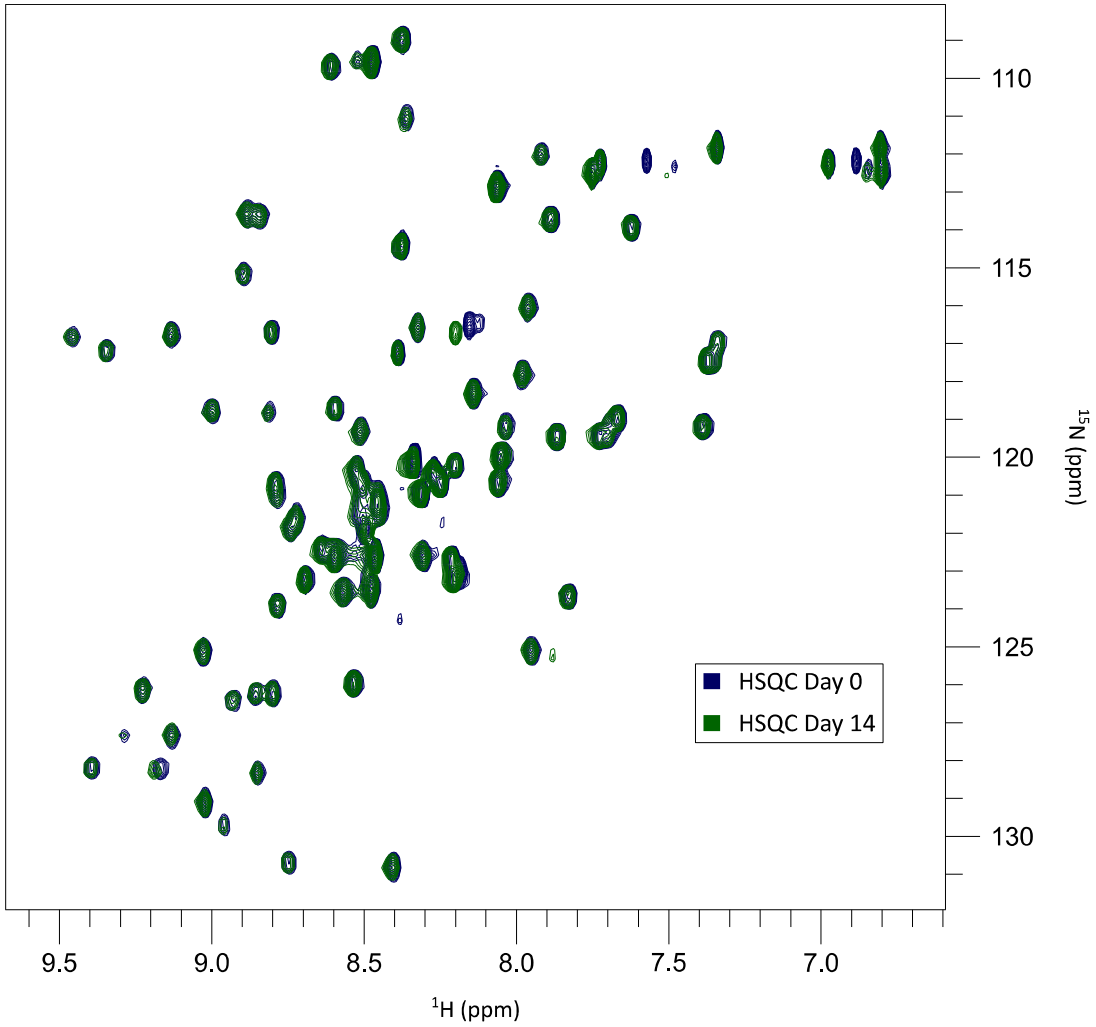


Figure 2.20: VWF C3 14-day stability test A comparison of a ^{15}N-HSQC of VWF C3 in 20 mM sodium phosphate, 50 mM NaCl, pH 7.4 performed on Day 0 and Day 14 at 298 K. The sample was left at room temperature for the intervening days to test its stability, with triple resonance experiments performed over multiple days. A peak shift was observed at peak 8.2 ppm (^{1}H)/116.7 ppm (^{15}N), however, minimal degradation was observed with a single sharp peak emerging at 7.9 ppm (^{1}H)/125.3 ppm (^{15}N).

2.8. Isotopically-labelled expression and NMR assignment completion

With purification, cleavage and buffer conditions optimised, a uniformly isotopically-labelled VWF C3 sample was expressed in two litres of M9 minimal media containing $^{15}NH_4Cl$ and $^{13}C_6$-D-glucose. This was once again purified by two-step Ni-NTA affinity and size exclusion chromatography (Figure 2.21). The expressed protein gave a final yield of 7 mg C3 protein / litre of culture, providing four samples at 900 µM for NMR analysis.

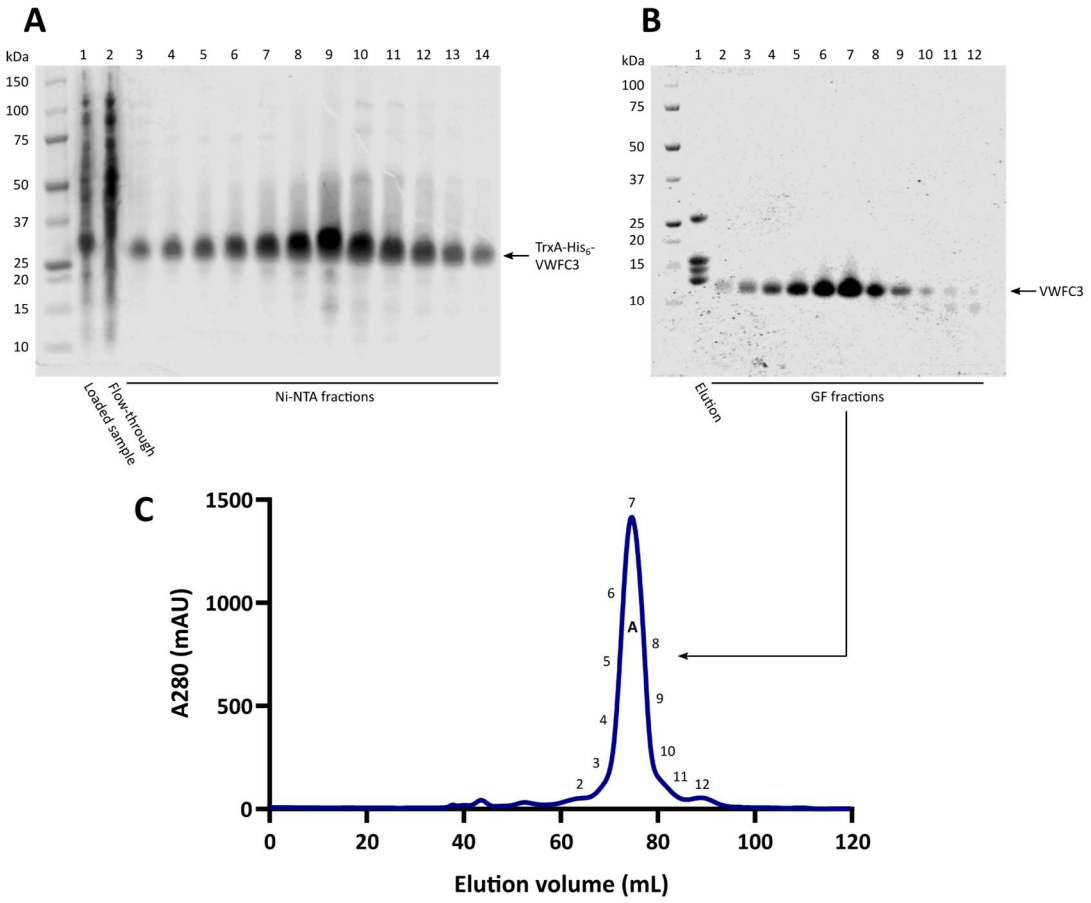


Figure 2.21: Purification of isotopically-labelled VWF C3 The Ni-NTA affinity chromatography purification of the uncleaved TrxA-His_6-VWFC3 is shown in Panel A, analysing the elution fractions in reducing SDS-PAGE. Lane 1 shows the contents extracted from the cell cytoplasm, with Lane 2 showing the unbound protein that flowed through the column. Fractions eluted from the column are shown in Lanes 3-14. Panel B analyses the elution fractions from the size exclusion chromatography (Panel C), following a reverse Ni-NTA. Lane 1 shows the bound protein that was subsequently eluted from the column, with the unbound protein in the flow-through was separated by size exclusion chromatography. Lanes 2-12 correspond to the numbers seen on the chromatogram in Panel C, with the large band at 13 kDa corresponding to cleaved VWF C3.

The equivalent triple resonance experiments were carried out as in Table 2.2 in order to check and complete the backbone amide assignment of VWF C3 (Figure 2.22). All assignments from VWF C3 (FXa) were transferred and checked, with the only assignment deemed incorrect being that of S2435, the peak of which was subsequently assigned to S2479.

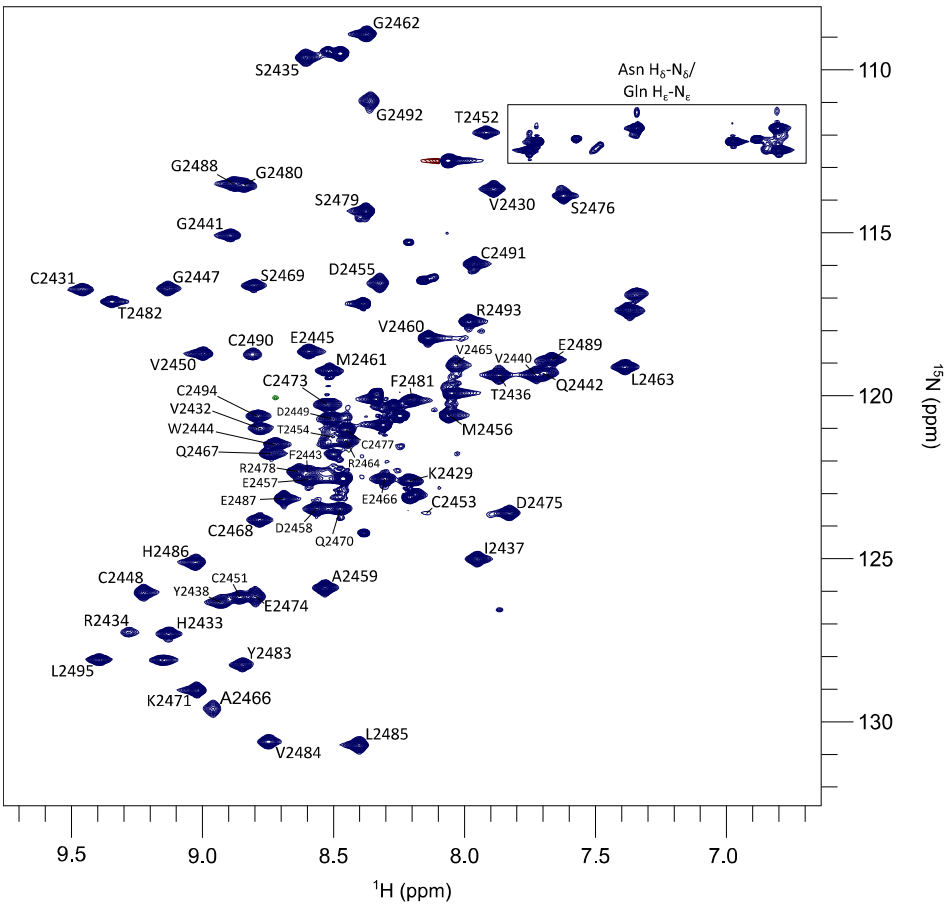


Figure 2.22: Backbone amide assignment of VWF C3 A full assignment of the backbone $^{1}H_{N}{}^{15}N_{H}$ of the VWF C3 domain. Unassigned peaks correspond to the linker impurity, present due to inaccurate cleavage.

The assignment also showed that inaccurate cleavage had once again occurred, with an assignment carried out from K2429 through the cleavage site (Asp-Asp-Asp-Asp-Lys) (Figure 2.23). Where this cleavage took place in the linker and why accurate cleavage did not occur remained unclear at the time (reason explained in Chapter 3.9). The only similar region where cleavage may have occurred is the Ala-Ala-Ala-Lys sequence, located 19 residues before K2429, however, due to the lack of apparent structure in the linker impurity causing the majority of peaks to be concentrated in the 8-8.5 ppm (^{1}H) range, these peaks are heavily overlapped, making an assignment of the N-terminal region extremely challenging. Whilst challenging, it appeared that the lack of structure in the linker impurity meant it was not incorporated and, therefore, interfering in the VWF C3 structure, likely not causing a problem in the structural assignment going forward. This would become clear during the assignment of the NOESY spectra (see Chapter 3.3), where NOEs are assigned to determine residues which are in close proximity and, hence, determine the secondary and tertiary folding of the structure. If the linker impurity presents NOEs to the VWF C3 domain, it may be incorporated into the structure and affecting the final calculations.

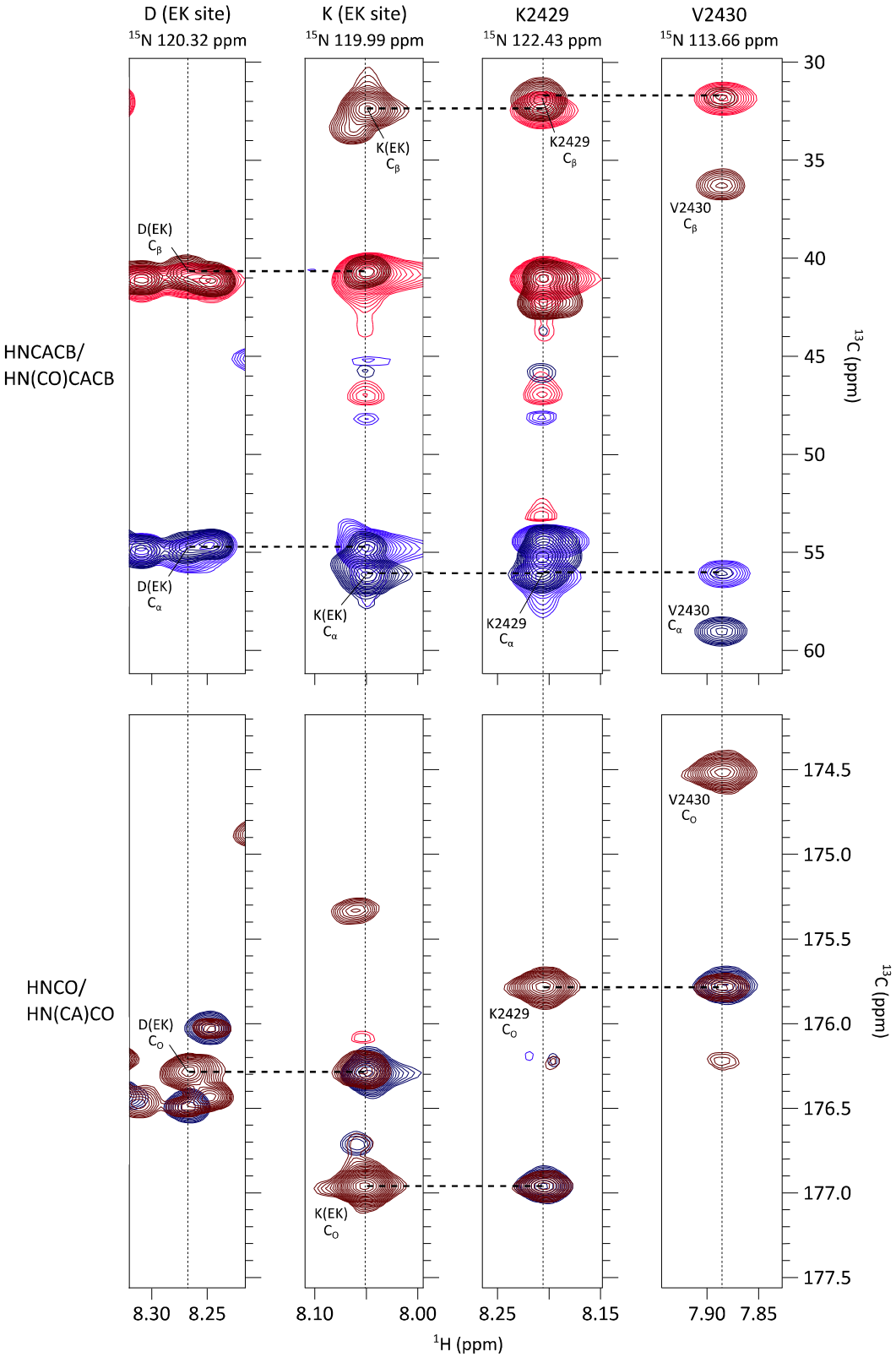


Figure 2.23: The assignment from K2429 to the EK cleavage site Shown is an assignment using both HNCACB/HN(CO)CACB and HNCO/HNC(CA)CO from V2430 to the aspartic acid of the EK cleavage site (D [EK site]). In the upper panel, $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ peaks from the HNCACB experiment are shown in a darker shade of blue and red, respectively. The lighter shade of blue and red represent the $^{13}C_{\alpha}^{i-1}$ and $^{13}C_{\beta}^{i-1}$, respectively, from the HN(CO)CACB experiment. In the bottom panel, blue peaks denote the $^{13}C_{O}^{i-1}$ from the HNCO experiment, with the red indicating the $^{13}C_{O}^{i}$ and $^{13}C_{O}^{i-1}$ from the HN(CA)CO. Sequential assignments between planes are indicated with dotted lines and assigned $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and $^{13}C_{O}$ of each residue is labelled.

3. Results: Three-dimensional structure of the VWF C3 domain

Whilst the assignment of the backbone ^{1}H-^{15}N nuclei, the "protein fingerprint", is the first key step on the path to determining a structure, assignment of the side chain nuclei chemical shifts, secondary structure and nuclei that are close in space, NOEs, is what provides information on the protein's tertiary structure. This assignment requires the use of triple resonance experiments, such as those seen in the backbone assignment, HNCACB and HN(CO)CACB. To identify neighbouring amino acid residues, triple resonance experiments rely upon spin-spin J-couplings to form a correlation between nuclei in residues *i* and *i-1*. Of particular importance are the one-bond coupling between $^{13}C_O^{i-1}$ and $^{15}N_H^{i}$ of ~15 Hz and the two-bond coupling of 4-9 Hz between $^{13}C_\alpha^{i-1}$ and $^{15}N_H^{i}$. The four-bond $^{1}H_\alpha^{i-1}$ and $^{1}H_N^{i}$ coupling, on the other hand, is far too small for such a correlation[410].

The chemical shift of a nucleus depends on the local magnetic field and, therefore, provides information on the surrounding chemical environment. This means a chemical shift profile can independently provide an insight into protein folding and predictions can be made on regions of secondary structure[411, 412]. Predictions, however, are used as a guide and by determining chemical shifts of aliphatic and aromatic side chains, we can ascertain the protein's secondary and tertiary structure. Assigning these chemical shifts and determining the protein structure is a starting point for progressing onto understanding a protein's dynamics and its function. This chapter presents the progression from the backbone amide assignment of VWF C3 to the final structural determination.

3.1. Assignment of the aliphatic ^{1}H and ^{13}C side chain chemical shifts

In order to conduct a complete structure calculation and determine the secondary and tertiary structure of a protein, assignment of the ^{13}C side chain chemical shifts is essential. These nuclei can be very important further down the line in the assignment of the NOESY spectra mentioned previously to determine the tertiary structure, with the side chains often forming hydrophobic/hydrophilic interactions, hydrogen and ionic bonds. Extensive coverage of the protein's chemical shift data is also vitally important to ensure accuracy of the structure calculations performed at a later date. The assignment of the $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts from the backbone assignment using the HNCACB and HN(CO)CACB experiments was transferred to begin the side chain assignment. Using this data with a number of other experiments (Table 3.1) allowed the identification of the side chain ^{13}C and ^{1}H chemical shifts. These assignments

were annotated on a ^{13}C-HSQC spectrum, a 2D spectrum providing a peak for each ^{1}H-^{13}C coupling.

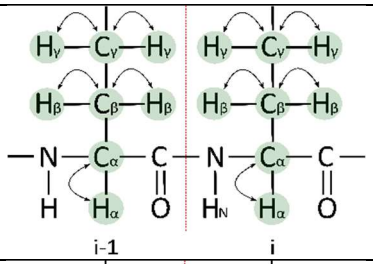
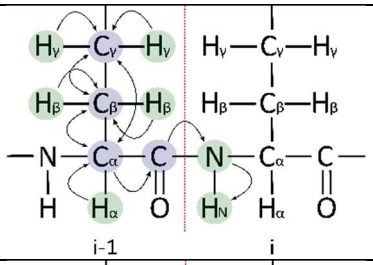
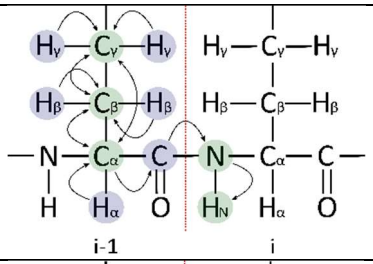
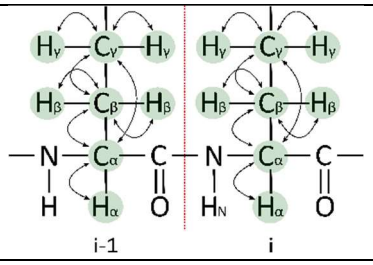
Experiment	Magnetisation transfer	Correlation
^{13}C-HSQC-TOCSY (2D)		$^{1}H_{\alpha}^{i}$-$^{13}C_{\alpha}^{i}$ $^{1}H_{\beta}^{i}$-$^{13}C_{\beta}^{i}$ $^{1}H_{\gamma}^{i}$-$^{13}C_{\gamma}^{i}$ $^{1}H_{\delta}^{i}$-$^{13}C_{\delta}^{i}$ Correlation shown for all ^{13}CH groups in the side chain
HCC(CO)NH-TOCSY (3D)		$^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$-$^{1}H_{\alpha}^{i-1}$ $^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$-$^{1}H_{\beta}^{i-1}$ $^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$-$^{1}H_{\gamma}^{i-1}$ $^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$-$^{1}H_{\delta}^{i-1}$ Correlation shown for all ^{13}CH groups in the side chain
CC(CO)NH-TOCSY (3D)		$^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$-$^{13}C_{\alpha}^{i-1}$ $^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$-$^{13}C_{\beta}^{i-1}$ $^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$-$^{13}C_{\gamma}^{i-1}$ $^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$-$^{13}C_{\delta}^{i-1}$ Correlation shown for all ^{13}CH groups in the side chain
HCCH-TOCSY		$^{1}H_{\alpha}^{i}$-$^{13}C_{\alpha}^{i}$ -$^{1}H_{\alpha}^{i}$/$^{1}H_{\beta}^{i}$/$^{1}H_{\gamma}^{i}$/$^{1}H_{\delta}^{i}$ $^{1}H_{\beta}^{i}$-$^{13}C_{\beta}^{i}$ -$^{1}H_{\alpha}^{i}$/$^{1}H_{\beta}^{i}$/$^{1}H_{\gamma}^{i}$/$^{1}H_{\delta}^{i}$ $^{1}H_{\gamma}^{i}$-$^{13}C_{\gamma}^{i}$ -$^{1}H_{\alpha}^{i}$/$^{1}H_{\beta}^{i}$/$^{1}H_{\gamma}^{i}$/$^{1}H_{\delta}^{i}$ $^{1}H_{\delta}^{i}$-$^{13}C_{\delta}^{i}$ -$^{1}H_{\alpha}^{i}$/$^{1}H_{\beta}^{i}$/$^{1}H_{\gamma}^{i}$/$^{1}H_{\delta}^{i}$ Correlation shown for all ^{13}CH groups in the side chain

Table 3.1: Experiments used in the side chain $^{1}H^{13}C$ chemical shift assignment The experiments used in the assignment of the backbone amide group, their magnetisation transfer and the resulting correlations observed. Letters in the experiment names correspond to the nuclei whose chemical shift is correlated, with those in brackets involved in the coherence transfer pathway. Nuclei highlighted in blue indicate where magnetisation is excited/transferred and green where the chemical shift is evolved. Note the CC(CO)NH-TOCSY experiment was used previously for the backbone $^{1}H_{N}^{15}N_{H}$ assignment.

The ^{13}C-HSQC experiment was performed as a constant-time HSQC which avoids the issue of peak splitting due to ^{13}C-^{13}C coupling and hence improves peak resolution. This experiment was performed with a reduced spectral width of 38 ppm, with the $^{13}C_{\alpha}$ region residing in the upper left quadrant. This method was also chosen to improve resolution, with a 38 ppm spectral width required in this case, rather than 60 ppm if it were to cover from the $^{13}C_{\delta}$ of I2437 (13 ppm) to the $^{13}C_{\beta}$ nuclei of threonine residues (up to 72 ppm). This amounts to a ~40% improvement in peak resolution in the ^{13}C dimension with the same number of indirect points acquired.

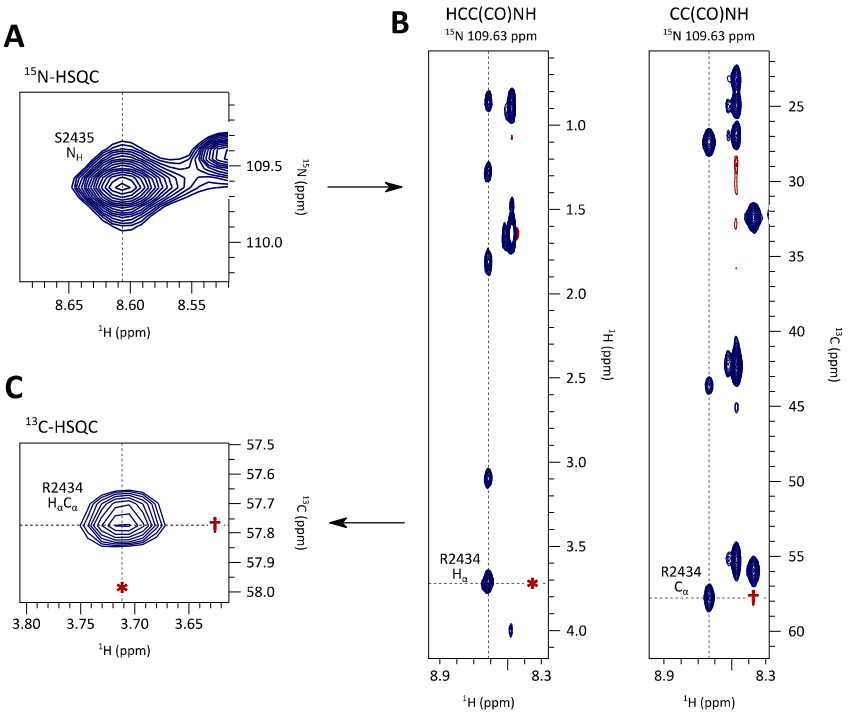


Figure 3.1: Transition from a backbone $^{1}H_{N}{}^{15}N_{H}$ assignment to a $^{1}H_{\alpha}{}^{13}C_{\alpha}$ assignment An example of the transition from the assignment of a backbone amide chemical shift to the $^{1}H_{\alpha}{}^{13}C_{\alpha}$ chemical shift using triple resonance experiments. Panel A shows the $^{1}H_{N}{}^{15}N_{H}$ assignment of S2435 on the ^{15}N-HSQC. The ^{1}H and ^{15}N chemical shifts shows the corresponding strips in the HCC(CO)NH and CC(CO)NH spectra, shown in Panel B. The ^{13}C chemical shift at the top of each strip corresponds to the chemical shift in the z-dimension. Each peak in the HCC(CO)NH strip corresponds to a ^{1}H chemical shift of a side chain ^{1}H nucleus in the preceding residue, R2434. The ^{13}C chemical shift at the top of each strip corresponds to the chemical shift in the z-dimension. Likewise, the strip in the CC(CO)NH spectrum gives the chemical shift of the ^{13}C side chain nuclei of R2434. Marked (* and †) are the $^{1}H_{\alpha}$ and $^{13}C_{\alpha}$ chemical shifts, respectively, of R2434 which correspond to the peak shown in the ^{13}C-HSQC (Panel C).

From a backbone amide assignment, the HCC(CO)NH and CC(CO)NH experiments displayed the chemical shifts of a residues aliphatic side chain ^{1}H and ^{13}C nuclei, respectively. Although $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts were already known from the backbone amide assignment using the HNCACB experiment, this method provided their corresponding $^{1}H_{\alpha}$ and $^{1}H_{\beta}$ chemical shift. Figure 3.1 outlines the transition from a backbone amide assignment to the $^{1}H_{\alpha}{}^{13}C_{\alpha}$ assignment with the remaining side chain nuclei observed in each strip (Panel B). In certain amino acids where each side chain nucleus has a significantly different chemical shift to its neighbours, such as alanine, tyrosine and arginine the side chain chemical shifts can be confidently assigned on the ^{13}C-HSQC. In the HCC(CO)NH strip shown in Figure 3.1 (Panel B), there are four peaks observed in the line of ~8.607 ppm, other than that already assigned to the $^{1}H_{\alpha}$ nucleus. The peak at ~3.1 ppm is that of the $^{1}H_{\delta}$, with the chemical shift of this nucleus typically isolated in this region. Closer inspection of the peak at ~1.8 ppm shows that this is two overlapped peaks, corresponding to the two $^{1}H_{\beta}$ nuclei. The remaining peaks at ~0.85 and ~1.3 ppm will, therefore, correspond to the two $^{1}H_{\gamma}$ nuclei. Likewise, in the CC(CO)NH strip, the peak at

~44 ppm is also a typical isolated peak of the $^{13}C_{\delta}$ nucleus. Closer inspection of the peak at ~27.5 ppm reveals two overlapped peaks, corresponding to the $^{13}C_{\beta}$ and $^{13}C_{\gamma}$ nuclei. Using the chemical shifts of their corresponding ^{1}H nuclei, we can determine the exact chemical shift of these nuclei by identifying cross-peaks observed in the ^{13}C-HSQC. The difference in peak strength of different nuclei also makes them easily distinguishable. For example, the $^{1}H_{\alpha}$ peak is noticeably stronger than $^{1}H_{\delta}$ due to the magnetisation transfer to $^{1}H_{N}$ occurring in four steps, rather than seven. With signal typically being lost at each transfer step, signal is better maintained in the case of $^{1}H_{\alpha}$ and results in a stronger peak. The same theory applies for the difference in signal strength between $^{13}C_{\alpha}$ and $^{13}C_{\delta}$ in the CC(CO)NH spectrum.

However, the majority of residues contain side chain nuclei with similar chemical shifts and less dispersion between peaks, meaning an assignment using these spectra and reference chemical shifts is less reliable. Furthermore, the HCC(CO)NH and CC(CO)NH often fail to detect nuclei further along the side chain from the $^{1}H_{N}$ nucleus due to the number of magnetisation transfer steps. This is particularly true for the $^{1}H_{\delta}/^{13}C_{\delta}$ nuclei and beyond. In these cases, an HCCH-TOCSY experiment was used in conjunction with the HCC(CO)NH and CC(CO)NH experiments. This was generally considered a more reliable side chain assignment method with detection occurring on each aliphatic ^{1}H nucleus, removing the possibility of loss of signal due to multiple magnetisation transfer steps as occurs in the HCC(CO)NH and CC(CO)NH. The HCCH-TOCSY plots the chemical shift of an aliphatic ^{1}H nucleus with a cross-peak to all ^{1}H nuclei in the aliphatic side chain, provided both ^{1}H nuclei are bound to a ^{13}C nucleus. The information provided by the HCC(CO)NH and CC(CO)NH is also insufficient for certain residues when used independently. In the case of valine, they do not provide information of which $^{1}H_{\gamma}$ chemical shift is correlated to which C_{γ} chemical shift, nor does it distinguish between nuclei with very similar chemical shifts in leucine ($^{1}H_{\beta}/^{1}H_{\gamma}$), proline ($^{1}H_{\beta}/^{1}H_{\gamma}$), arginine ($^{1}H_{\beta}/^{1}H_{\gamma}$) and lysine ($^{13}C_{\beta}/^{13}C_{\gamma}$) As previously mentioned, the benefit of using a TOCSY rather than COSY experiment is that cross-peaks are displayed for all side chain nuclei. This is an advantage with being able to identify similarities in peak patterns between different regions of the spectrum and helps to confirm peaks that originate from the same side chain as each other. Figure 3.2 is an example of the assignment of the ^{1}H and ^{13}C aliphatic side chain chemical shifts of I2437 using the HCCH-TOCSY.

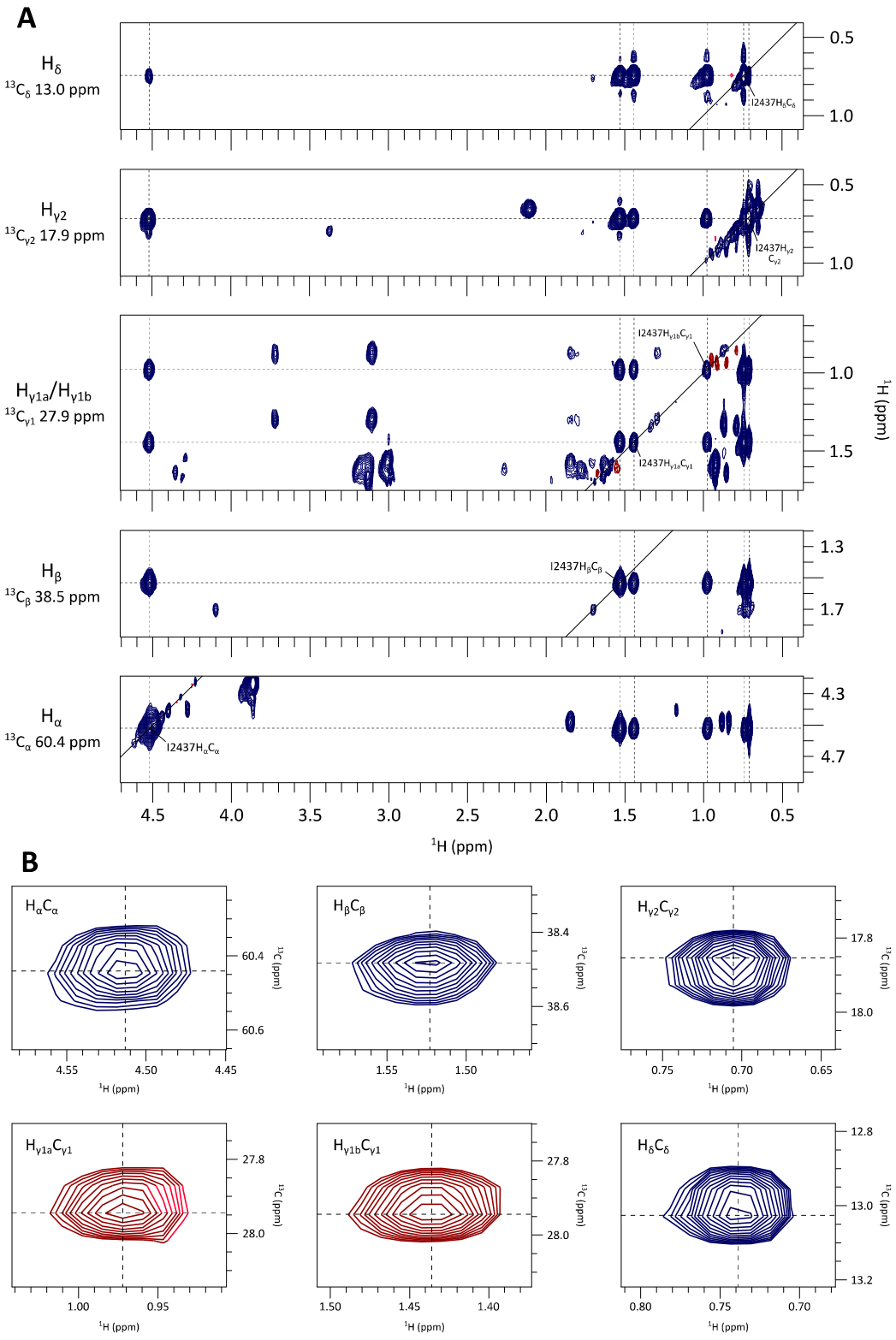


Figure 3.2: Assignment of the aliphatic ^{1}H and ^{13}C nuclei of I2437 using the HCCH-TOCSY experiment
Panel A shows the assignment of the side chain nuclei of I2437 using the HCCH-TOCSY whilst Panel B displays the assigned peaks of I2437 in the ^{13}C-HSQC. Each strip in Panel A corresponds to a ^{13}C nucleus showing the cross-peaks for all neighbouring nuclei. The lower strip of Panel A, at the $^{13}C_{\alpha}$ chemical shift of 60.4 ppm, presents a peak at ~4.5 ppm corresponding to the $^{1}H_{\alpha}$ nucleus. This data corresponds to the $^{1}H_{\alpha}{}^{13}C_{\alpha}$ peak in the ^{13}C-HSQC, upper left peak in Panel B. The same concept applies to the other six assigned peaks. Blue peaks correspond to a positive peak and red corresponds to a negative peak. The $^{1}H_{\gamma1}{}^{13}C_{\gamma1}$ peaks are negative as $^{13}C_{\gamma1}$ has an odd number of directly bound ^{13}C atoms. The ^{13}C chemical shift at the left of each strip corresponds to the chemical shift in the z-dimension.

Using $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts determined from the backbone assignment can be a starting point in the aliphatic side chain assignment, including proline residues. By identifying strips of identical peak arrangement at the two ^{13}C chemical shifts, ^{1}H chemical shifts of other aliphatic nuclei can be identified. For example, in the $^{1}H_{\alpha}$ and $^{1}H_{\beta}$ strips of I2437 (Figure 3.2, Panel A), the similarities in the peak positions in the direct ^{1}H dimension is immediately noticeable, with both presenting six peaks at identical ^{1}H chemical shifts. This provides the ^{1}H chemical shifts of the remainder of the side chain nuclei and by scanning through the spectrum along the ^{13}C dimension with the CC(CO)NH spectrum as a guide, we can identify the same pattern of ^{1}H peaks at a ^{13}C chemical shift of 27.9 ppm, 17.9 ppm and 13 ppm. The reference chemical shifts of the isoleucine ^{13}C nuclei can determine which strip corresponds to which ^{13}C nucleus, with these mentioned corresponding to $^{13}C_{\gamma 1}$, $^{13}C_{\gamma 2}$ and $^{13}C_{\delta}$ respectively. In each strip, the diagonal peak with identical ^{1}H chemical shift in both the direct and indirect ^{1}H dimensions corresponds to the chemical shift of the respective attached ^{1}H nucleus. The assignment by the HCCH-TOCSY method can be used conjunction with that observed in the HCC(CO)NH and CC(CO)NH by means of confirmation.

For the stereospecific assignment of the methyl groups of valine ($^{1}H_{\gamma}{}^{13}C_{\gamma}$) and leucine ($^{1}H_{\delta}{}^{13}C_{\delta}$) residues, a VWF C3 sample was expressed in SHuffle *E. coli* cells[400] in M9 minimal media as in Chapter 2.7, though using 10% $^{13}C_6$-D-glucose and 90% $^{12}C_6$-$^{1}H_{12}$-glucose, known as fractional labelling (Figure 1.19)[389]. A constant-time ^{13}C-HSQC was performed on this 10% ^{13}C-labelled C3 sample with peaks corresponding to Pro-R and Pro-S ^{13}C nuclei residing in opposite phases. Processing of the spectrum meant that Pro-R methyl groups were negative with an even number of bound ^{13}C nuclei ($^{13}C_{\gamma 1}$-$^{13}C_{\beta}$ in valine and $^{13}C_{\delta 1}$-$^{13}C_{\gamma}$ in leucine), conversely, Pro-S methyl groups were positive with an odd number of ^{13}C nuclei directly bound ($^{13}C_{\gamma 2}$ in valine and $^{13}C_{\delta 2}$ in leucine) (Figure 3.3). This difference in phasing due to the ^{13}C-^{13}C coupling between two ^{13}C-labelled nuclei is the same concept as that displayed in the HCCH-TOCSY of I2437 (Figure 3.2, Panel B) where the peaks from the $^{13}C_{\gamma 1}$ nucleus are in a negative phase due to an odd number of directly bound ^{13}C nuclei.

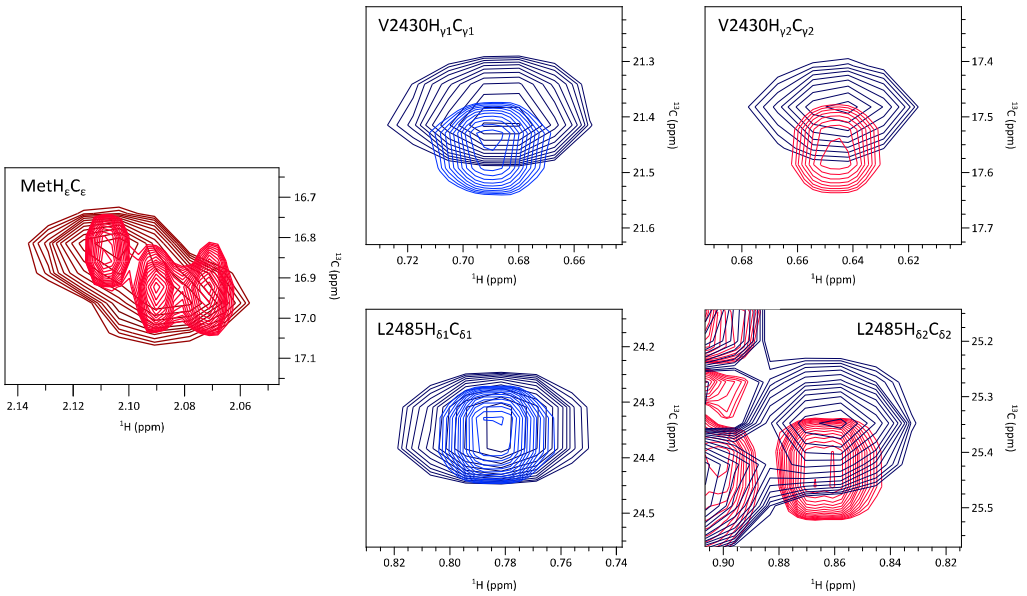


Figure 3.3: Stereospecific assignment of the V2430 and L2485 side chain nuclei using fractional labelling An assignment of the γ1 and γ2 methyl groups of V2430 and δ1 and δ2 methyl groups of L2485. Peaks in dark blue and red correspond to the positive and negative peaks, respectively, from the constant time ^{13}C-HSQC of the 100% ^{13}C-labelled C3 sample. Peaks in light blue and red correspond to the positive and negative peaks, respectively, from the constant time ^{13}C-HSQC of the 10% ^{13}C-labelled C3 sample. The far-left window (MetHεCε) shows the peaks corresponding to the methyl group of MetHεCε, the assignment of which was carried out at a later date. Other windows indicate the stereospecific assignment of the methyl groups of V2430 (above) and L2485 (below).

The positive/negative phasing was determined by the distinct peaks of MetHεCε. These peaks are well separated in the ^{13}C-HSQC spectrum from peaks of any other side chain nuclei, with these being the sole peaks greater than 2 ppm in the ^{1}H dimension at a ^{13}C chemical shift of ~17 ppm. With MetHεCε separated from the remainder of the side chain by a sulphur atom, this atom when enriched by ^{13}C it is guaranteed to not be bound to a second ^{13}C atom. Therefore, this peak will display the same sign as that of the $^{1}H_{\gamma2}{}^{13}C_{\gamma2}$ of valine or $^{1}H_{\delta2}{}^{13}C_{\delta2}$ of leucine with these being bound to a ^{12}C atom in 90% of cases. The $^{1}H_{\gamma1}{}^{13}C_{\gamma1}$ of valine or $^{1}H_{\delta1}{}^{13}C_{\delta1}$ of leucine, however, will display the opposite sign, with it being bound to a ^{13}C-labelled atom in each case of itself being enriched with ^{13}C (Figure 3.3).

This completed a 98% assignment of the backbone $^{1}H_N$,$^{15}N_H$ and $^{13}C_O$ nuclei and ^{1}H and ^{13}C nuclei of the backbone ($^{1}H_\alpha{}^{13}C_\alpha$) (Figure 3.4) and aliphatic side chains (662 of 675 nuclei), with those left unassigned detailed in Table 3.2. Note at this stage, this assignment excludes the aromatic side chain assignments of phenylalanine, histidine, tryptophan and tyrosine, the side chain amide groups of glutamine and the $^{1}H_\varepsilon{}^{13}C_\varepsilon$ of methionine residues. These were assigned at a later date using NOESY spectra.

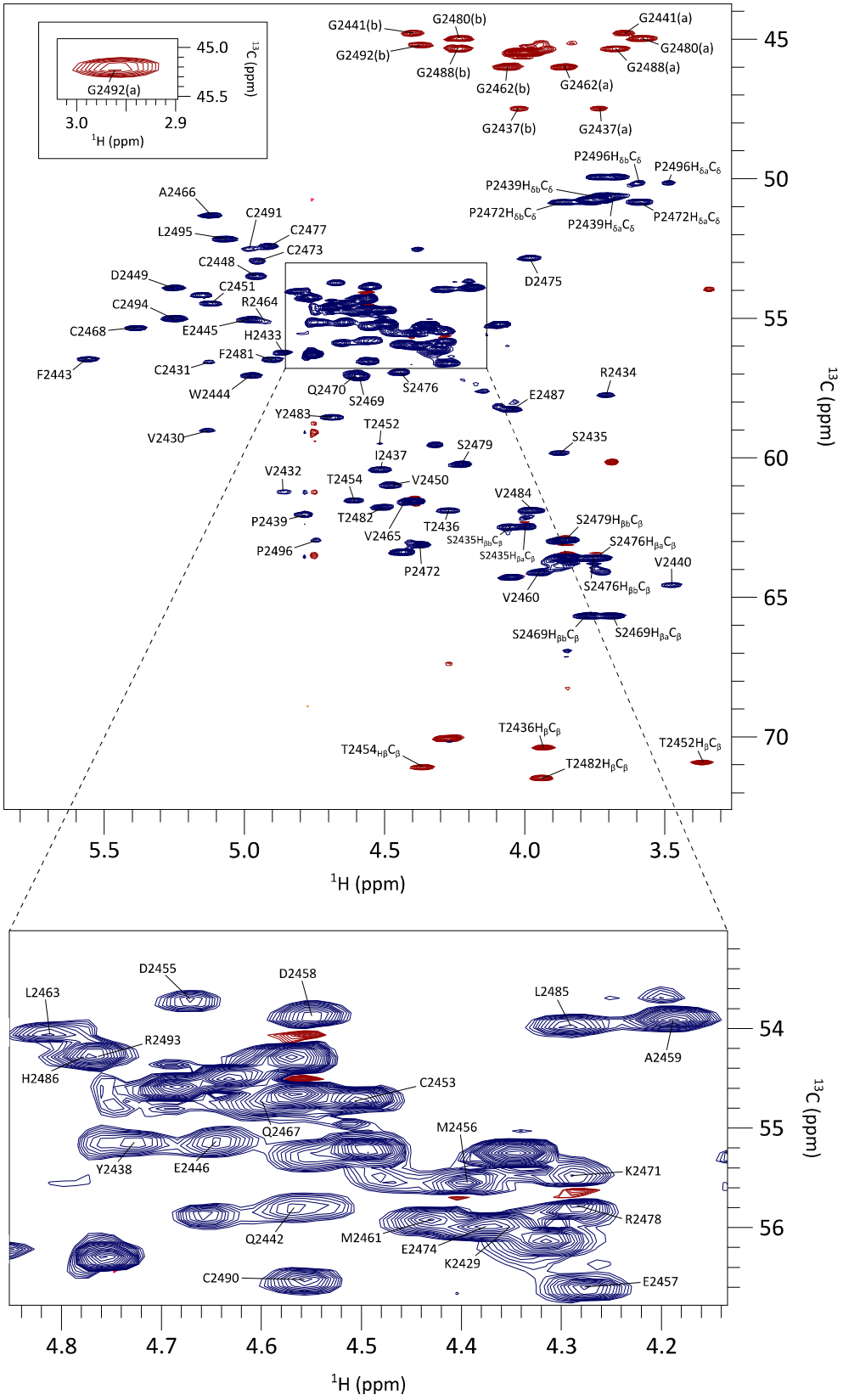


Figure 3.4: ^{13}C-HSQC assignment of the VWF C3 $^{13}C_{\alpha}$ region A constant time ^{13}C-HSQC region showing the assignment of the $^{1}H_{\alpha}{}^{13}C_{\alpha}$ chemical shifts of VWF C3 with positive (blue) and negative (red) peaks. Positive peaks arise from an even number of directly bound ^{13}C atoms while an odd number of directly bound ^{13}C atoms results in a negative peak. Labels correspond to the $^{1}H_{\alpha}{}^{13}C_{\alpha}$ assignment unless otherwise stated. A zoom of the central region is presented in the bottom panel and $^{1}H_{\alpha a}{}^{13}C_{\alpha}$ of G2492, beyond the limits of the displayed spectrum, is shown in the top left of the upper panel.

Residue	Unassigned nuclei
K2429	$^{1}H_{\varepsilon b}$
R2434	$^{1}H_{\delta b}$
E2446	$^{1}H_{\gamma b}$
K2471	$^{1}H_{\delta b}$ $^{1}H_{\varepsilon b}$
R2478	$^{1}H_{\gamma b}$ $^{1}H_{\delta b}$
E2487	$^{1}H_{\gamma b}$
E2489	$^{1}H_{\gamma b}$

Table 3.2: Unassigned ^{1}H nuclei of VWF C3 This excludes aromatic side chains, side chain amide groups and $^{1}H_{\varepsilon}$ of methionine residues which were assigned using NOESY spectra.

3.2. Secondary structure assignment

A TALOS+ program was used for a number of predictions including backbone rigidity, phi and psi backbone torsion angles and secondary structure (Figure 3.5). It uses the chemical shift data from the $^{1}H_{N}$, $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C_{O}$ and $^{15}N_{H}$ nuclei and makes predictions using a neural network classification scheme and a database of 200 proteins. The TALOS+ approach involves searching a high resolution database for the ten best matches to the chemical shift of a given residue in the protein of interest. If there is a consensus of phi and psi angles, this forms a prediction for the backbone angles of that residue. It also analyses the chemical shift and sequence to predict whether a given residue being is in a sheet, helix or loop conformation with a prediction accuracy of approximately 89%. In order to determine secondary structure, and subsequently tertiary structure, three NOESY experiments were carried out (Table 3.3), providing NOEs which display nuclei which are close in space but not necessarily chemically bound. NOESY spectra only display NOEs for protons which are no more than 6 Å apart.

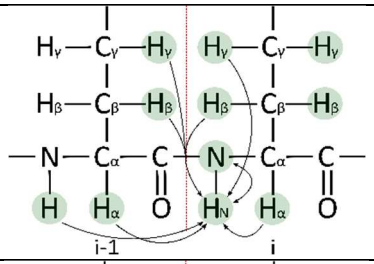
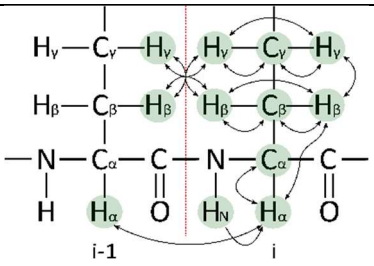
Experiment	Magnetisation transfer	Correlation
^{15}N-edited NOESY (3D)		$^{1}H_{N}$-$^{15}N_{H}$-x where x is any neighbouring ^{1}H nucleus (Occurs for any backbone or side chain NH group)
^{13}C-edited NOESY (3D) (Carried out on the aliphatic and aromatic ^{13}C regions)		$^{1}H_{\alpha}$-$^{13}C_{\alpha}$-x $^{1}H_{\beta}$-$^{13}C_{\beta}$-x $^{1}H_{\gamma}$-$^{13}C_{\gamma}$-x where x is any neighbouring ^{1}H nucleus (Occurs for any aliphatic or aromatic ^{13}CH group)

Table 3.3: 3D NOESY experiments for the secondary and tertiary structural assignment The two types of NOESY experiments performed for the structural assignment. Two types of ^{13}C-edited NOESY experiments were carried out, one with a ^{13}C offset in the aliphatic region and one in the aromatic region. Each diagonal peak displays cross-peaks to neighbouring ^{1}H nuclei.

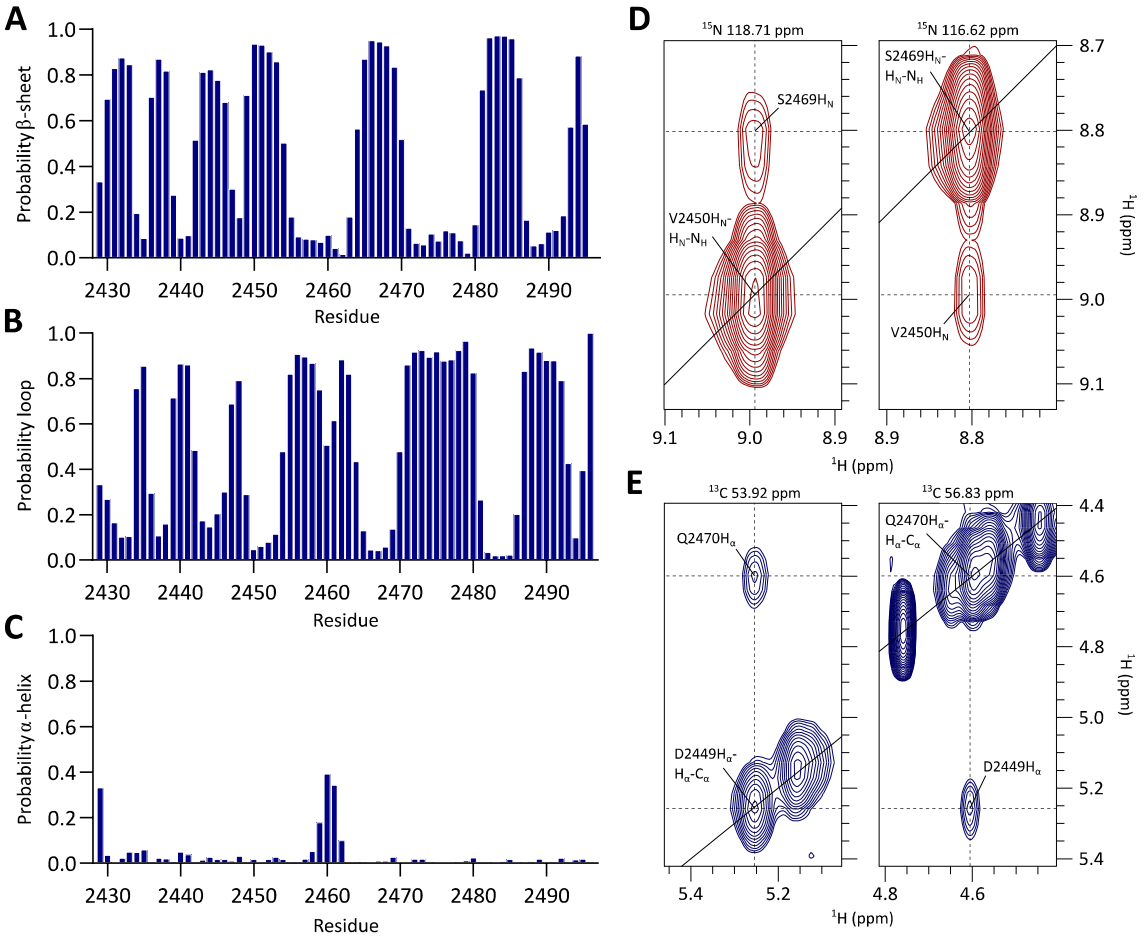


Figure 3.5: TALOS+ secondary structure prediction and NOE assignment Displayed in Panels A, B and C is the TALOS+ prediction of regions of β-sheet, loop and α-helix, respectively. Panels D and E display a NOE assignment between two strands of β-sheet. Panel D uses a ^{15}N-edited NOESY experiment, with the left panel showing a diagonal peak corresponding to V2450^{1}H$_N$ and a cross-peak to S2469^{1}H$_N$. The right panel shows a diagonal peak corresponding to S2469^{1}H$_N$ and a cross-peak in reverse to V2450^{1}H$_N$. Panel E uses a ^{13}C-edited NOESY to display a NOE between D2449^{1}Hα and Q2470^{1}Hα. Both of these NOE assignments are between the same two β-sheet strands. The ^{15}N/^{13}C chemical shift at the top of each strip corresponds to the chemical shift in the z-dimension.

The TALOS+ prediction suggests that the VWF C3 structure is dominated by β-sheet and loop regions (Figure 3.5, Panels A and B). Using the chemical shift data, it predicted seven strands of β-sheet, as expected from other VWFC domain structures. All intervening regions between β-sheet strands has a high degree of probability of loop structures, with the α-helix probability across the whole protein being extremely low. The highest probability for α-helix was at residues 31, 32 and 33 (A2459, V2460 and M2461) which had a probability of 0.18, 0.393 and 0.343, respectively, however, the probability of a loop at these three residues was far higher at 0.751, 0.507 and 0.615, respectively. The probability of α-helix in the remaining residues was no higher than 0.1. Whilst not a precise determination of secondary structure, particularly their N- and C-terminal boundaries, the prediction gave guidance as to the likely location of β-sheet strands and the regions to focus on for the β-sheet assignment in the NOESY spectra.

NOESY spectra are characterised by a series of diagonal peaks corresponding to a nucleus with a series of cross-peaks, or NOEs, in the direct ^{1}H dimension, aligning at the ^{1}H chemical shift of that nucleus in the indirect dimension. At the ^{15}N chemical shift of that proton's attached ^{15}N nucleus, a diagonal peak presents a cross-peak in reverse. For example, the ^{15}N-edited NOESY at a ^{1}H chemical shift of 8.99 ppm and a ^{15}N chemical shift of 118.71 ppm presents a diagonal peak corresponding to V2450$^{1}H_{N}$ (Figure 3.5, Panel D). This presents a cross-peak to S2469$^{1}H_{N}$ at 8.81 ppm. At the $^{15}N_{H}$ chemical shift of S2469, 116.62 ppm, a diagonal peak corresponding to the S2469$^{1}H_{N}$ nucleus is observed at 8.81 ppm. This presents a cross-peak in reverse to V2450$^{1}H_{N}$ (Figure 3.5), allowing a NOE assignment between these two nuclei. The ^{13}C-edited NOESY assignment works in the same manner, with peaks correlated with their attached ^{13}C nucleus rather than ^{15}N. Two β-sheet strands organise in an ordered manner with NOEs seen between backbone amide groups (Figure 3.5, Panel D) and between $^{1}H_{\alpha}$ nuclei (Panel E). Displayed in Figure 3.6 are the NOE assignments for the three assigned β-sheets.

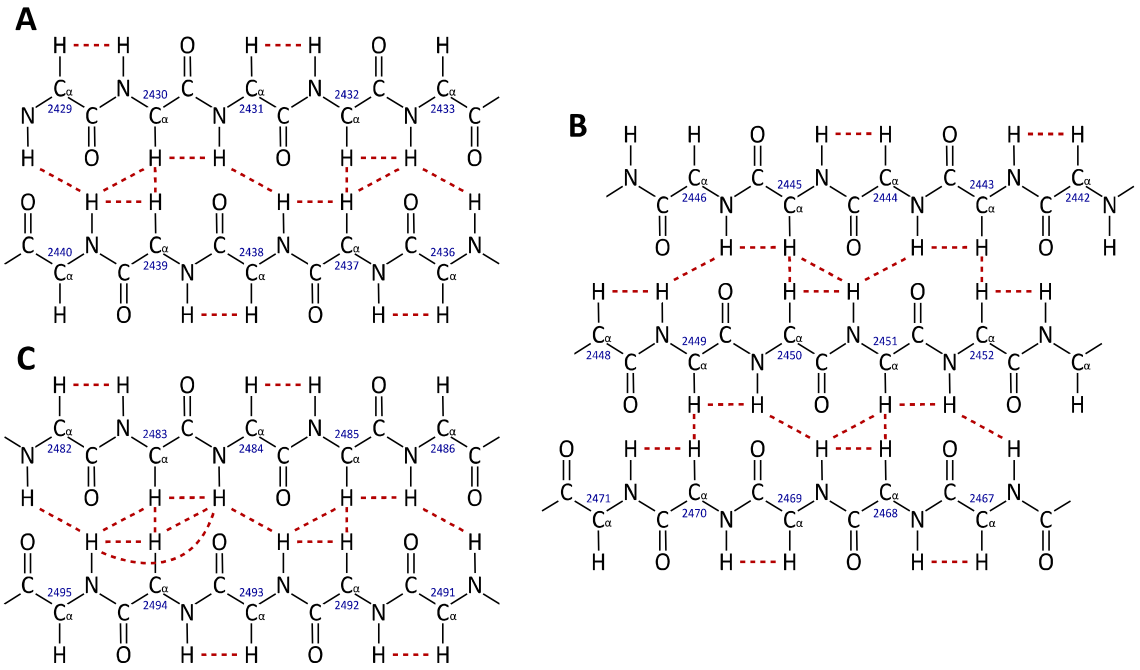


Figure 3.6: Assignment of β-sheet NOEs Displayed is the assignment of the three VWF C3 antiparallel β-sheets. A red dotted line indicates a NOE assignment while blue numbers indicate residue number. Panel A shows a β-sheet between strands of K2429 to H2433 and T2436 to V2440. Panel B shows a triple stranded β-sheet between strands of residues Q2442 to E2446, C2448 to T2452 and Q2467 to K2471. Panel C shows the final β-sheet between strands of T2482 to H2486 and C2491 to L2495.

The NOE assignment (Figure 3.6) displays the organised manner which β-sheets organised into. Not only are inter-strand NOEs observed but also intra-strand NOEs between adjacent $^{1}H_{N}$ and $^{1}H_{\alpha}$ nuclei. The result of this secondary structure assignment is, once again, as expected with two double-stranded antiparallel β-sheets and one triple-stranded antiparallel β-sheet, similar

to what is observed in other VWFC domains. This also provides an insight into one disulphide bond in VWF C3, with C2451 and C2468 in close proximity in the triple-stranded β-sheet, suggesting that the interaction of these two β-sheet strands is strengthened by the presence of a disulphide bond.

3.3. Disulphide bonds and tertiary structure assignment

The assignment of the VWF C3 disulphide bonds is essential to accurately determining its tertiary structure. In order to determine whether the disulphide bonds had formed, an Ellman's assay was carried out which determines the concentration of free thiols in a protein sample. By using a series of standards which had a free thiol concentration ranging from 0 mM to 1.5 mM in a series of 0.25 mM intervals and observing their absorption at 412 nm, a standard curve was produced. The absorbance of a VWF C3 sample at 412 nm, meanwhile, was recorded as 0.002. With the equation determined from the cysteine standards indicating a free thiol concentration in the C3 sample of -2.1 μM (Table 3.4), it suggested that all disulphide bonds had formed.

Standard	Free thiol concentration (mM)	OD_{412}
A	1.5	1.901
B	1.25	1.598
C	1	1.268
D	0.75	0.949
E	0.5	0.630
G	0.25	0.312
F	0	0
VWF C3	**-0.0021**	**0.003**

Table 3.4: Ellman's Assay for the free thiol concentration of the C3 protein Displayed is the Ellman's Assay result, showing the free thiol concentration and OD_{412} values of each standard (A-F). Displayed in bold is the free thiol concentration of the expressed C3 protein, calculated using the trendline equation of the standards, $y = 1.2733x + 0.0038$, which had an R^2 value of 0.9999.

The $^{13}C_\beta$ chemical shift of cysteine residues can also predict whether those cysteines are disulphide bonded. These chemical shift values were determined during the backbone amide assignment using the HNCACB experiment.

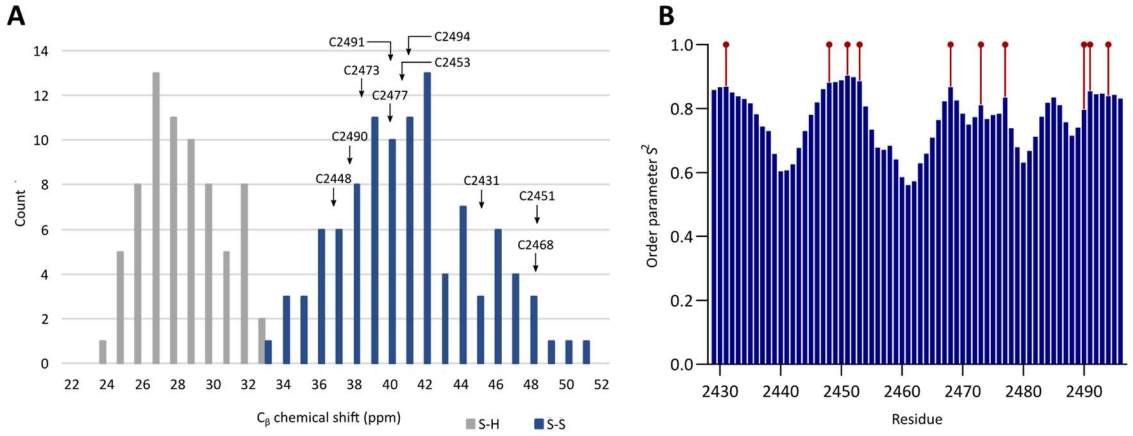


Figure 3.7: Chemical shift variations of cystine $^{13}C_\beta$ nuclei and VWF C3 backbone rigidity Panel A displays the distribution plot of cysteine $^{13}C_\beta$ chemical shifts when the cysteines are reduced, grey bars, and oxidised, blue bars. Reduced cysteines have a $^{13}C_\beta$ chemical shift between 24 and 33 ppm, whilst oxidised cysteines have a $^{13}C_\beta$ chemical shift between 33 ppm and 51 ppm. This was derived by Sharma et al[413]. Marked by the arrows above are the $^{13}C_\beta$ chemical shifts of the VWF C3 cysteine residues. All reside in the region which suggests they are disulphide bonded. A prediction of backbone rigidity from TALOS+ is shown in Panel B in the form of order parameter S^2, with 0 denoting isotropic motion and 1 indicating complete rigidity. Marked in red are the VWF C3 cysteine residues.

Both the Ellman's assay data and the $^{13}C_\beta$ chemical shifts of cysteine residues (Figure 3.7, Panel A) suggest that all ten cysteines are disulphide bonded. The TALOS+ prediction of backbone rigidity (Figure 3.7, Panel B) by order parameter S^2 uses chemical shift data to display fluctuations of a backbone N-H bond vector due to its internal motion[414]. Noticeably, the most rigid regions are centred around the disulphide bonds except for the region centred around L2485, though residues T2482 to H2486 form one strand of a β-sheet, adding to its rigidity. All predicted loop regions (Figure 3.5, Panel B) tend to correspond to the most flexible regions except for the region from P2472 to C2477. Despite being a loop region, it is stabilised by two disulphide-bonded cysteines (C2473 and C2477) flanking this region. This would suggest that these cysteines are not bonded to each other as such a bond would not add flexibility to this loop region. These results highlight the importance of the VWF C3 disulphide bonds in forming a stable tertiary structure with a high degree of flexibility in the intervening regions between disulphide bonds.

Once disulphide bond formation was confirmed, determining which cysteines these formed between was essential. This was performed using the NOESY spectra. Whilst a NOE between two sulphur atoms is not possible, a NOE assignment ideally between two $^1H_\beta$ cysteine nuclei

indicates a NOE between those residues. A clear assignment between these nuclei, however, may not always be possible with these having very similar ^{1}H chemical shifts to each other, causing overlap of diagonal and cross-peaks. A NOE assignment can be performed between two other nuclei in those residues, the $^{13}C_{\alpha}$ nuclei for example, or between one cysteine and the neighbouring residue of its reciprocating cysteine. Whilst these give an initial idea of the disulphide bond arrangement, these can be confirmed during the tertiary structure assignment with NOEs providing a clearer picture of the domain's overall folding.

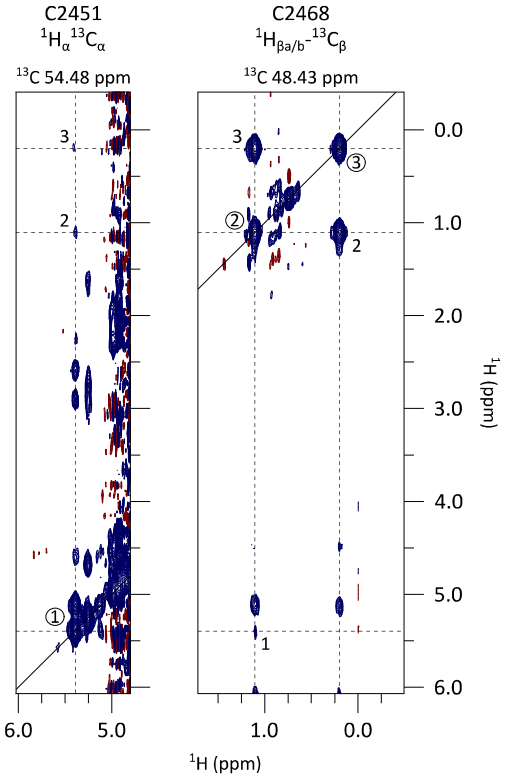


Figure 3.8: NOE assignment between the $^{1}H_{\alpha}$-$^{1}H_{\beta}$ nuclei of C2451 and C2468 The disulphide bond assignment between C2451 and C2468, with a NOE assignment between the C2451$^{1}H_{\alpha}$ and C2468$^{1}H_{\beta}$ nuclei. Blue peaks are positive peaks whilst red peaks are negative. The nuclei labelled above each strip corresponds to the diagonal peak in that strip. Each diagonal peak is labelled with a circled number and each cross-peak labelled with a number corresponding to the diagonal peak it forms a NOE with. Note a cross-peak from C2468$^{1}H_{\beta b}$ nucleus can be observed when zoomed into the baseline noise. The ^{13}C chemical shift at the top of each strip corresponds to the chemical shift in the z-dimension.

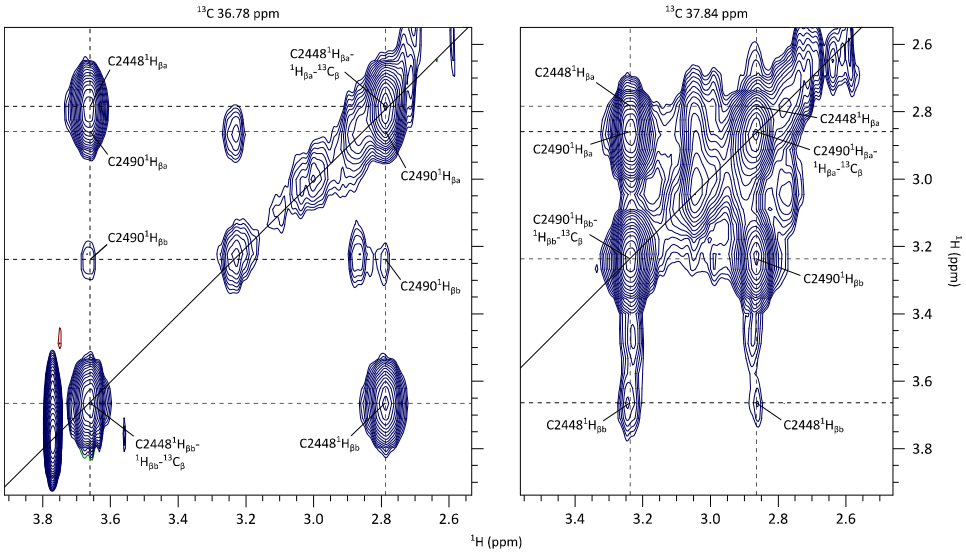


Figure 3.9: NOE assignment between $^1H_\beta$ nuclei of C2448 and C2490 This uses the ^{13}C-edited NOESY to assign the NOE between the two $^1H_\beta$ nuclei of C2448 and C2490, indicating that these two residues are disulphide bonded. The left panel, at the $^{13}C_\beta$ chemical shift of C2448, shows a diagonal peak for each of the two $^1H_\beta$ nuclei and cross-peaks to the two $^1H_\beta$ nuclei of C2490. The right-hand panel, at the $^{13}C_\beta$ chemical shift of C2490, shows this NOE in reverse with a cross-peak to the two $^1H_\beta$ nuclei of C2448. The ^{13}C chemical shift at the top of each strip corresponds to the chemical shift in the z-dimension.

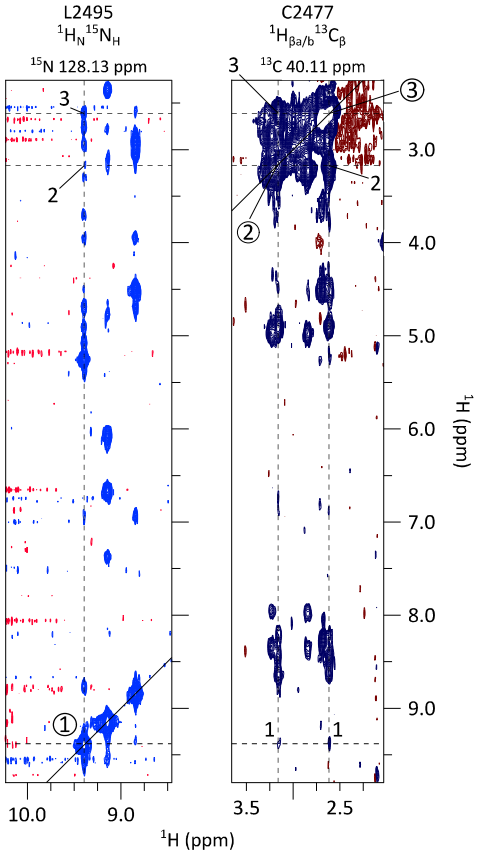


Figure 3.10: NOE assignment indicating a C2477-C2494 disulphide bond This displays a NOE between the $^1H_\beta$ nuclei of C2477 and the ^1H_N nucleus of L2495, the neighbouring residue to C2494. The left panel shows a ^{15}N-edited NOESY and the right-hand panel a ^{13}C-edited NOESY. Blue peaks are positive whilst red peaks are negative. Each diagonal peak is labelled with a circled number and each cross-peak labelled with a number corresponding to the diagonal peak it forms a NOE with. The ^{15}N/^{13}C chemical shift at the top of each strip corresponds to the chemical shift in the z-dimension.

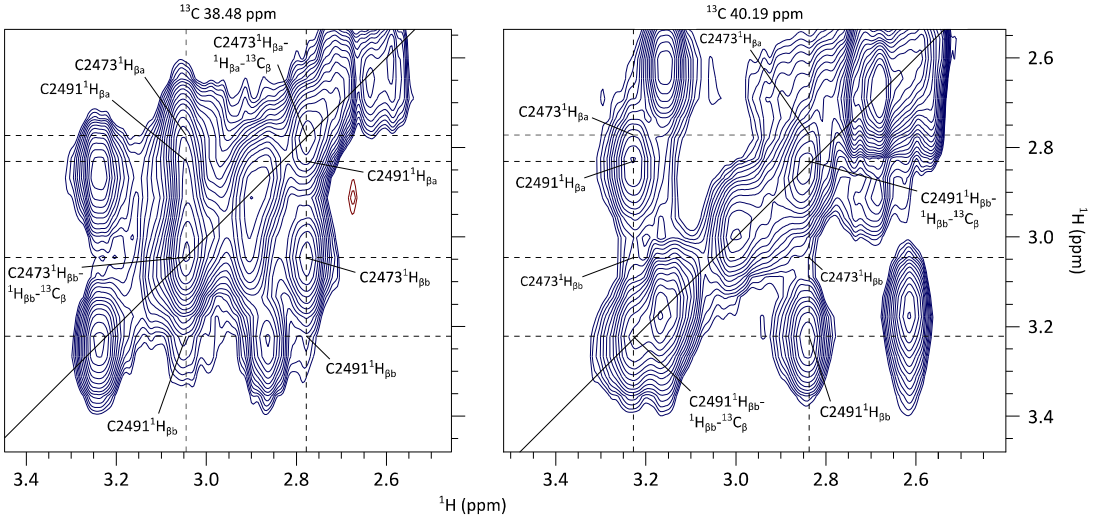


Figure 3.11: NOE assignment between the $^{1}H_{\beta}$ nuclei of C2473 and C2491 This NOE assignment indicates a disulphide bond between C2473 and C2491. The left panel shows, at the $^{13}C_{\beta}$ chemical shift of C2473, two diagonal peaks corresponding to the two $^{1}H_{\beta}$ nuclei of C2473. This panel also displays two cross-peaks to the $^{1}H_{\beta}$ nuclei of C2491. The right-hand panel shows this NOE assignment in reverse with a cross-peak to the C2473$^{1}H_{\beta}$ nuclei observed at the $^{13}C_{\beta}$ chemical shift of C2491. The ^{13}C chemical shift at the top of each strip corresponds to the chemical shift in the z-dimension.

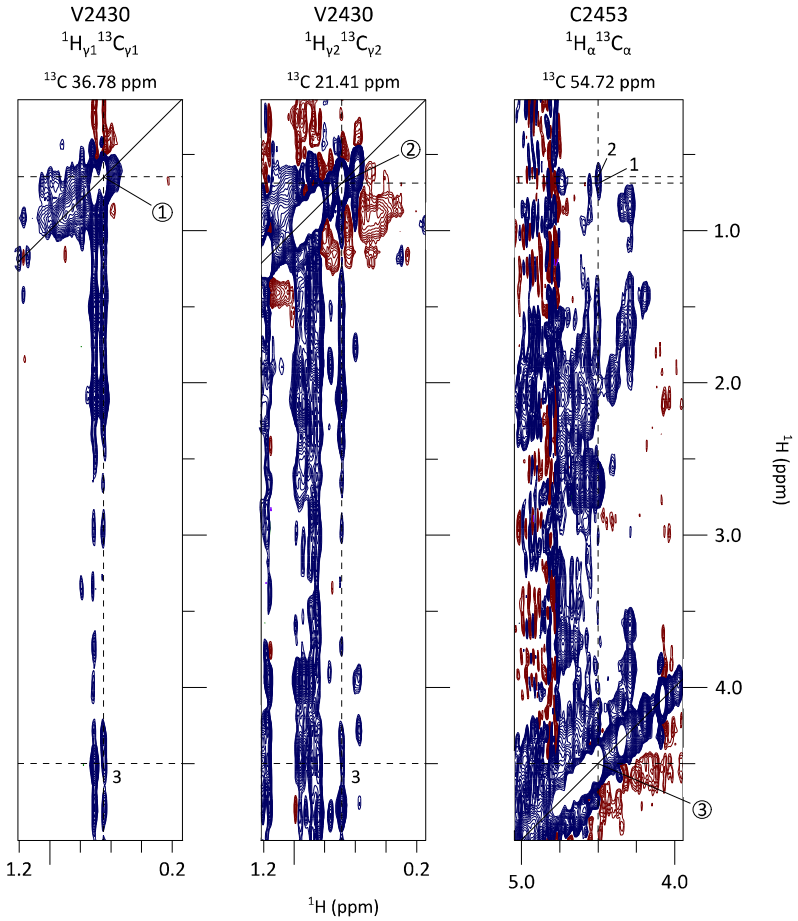


Figure 3.12: A NOE assignment indicating a C2431-C2453 disulphide bond A NOE assignment between V2430 and C2453 suggests a disulphide bond between C2453 and the neighbouring residue to V2430, C2431. The blue peaks are positive peaks and the red peaks are negative. Each diagonal peak is labelled with a circled number and each cross-peak labelled with the number of the diagonal peak it forms a NOE with. The ^{13}C chemical shift at the top of each strip corresponds to the chemical shift in the z-dimension.

The NOE assignments (Figure 3.8 to 3.12) indicate a full disulphide bond assignment, shown below. This arrangement matches that predicted by sequence analysis of the VWF C domain region[37].

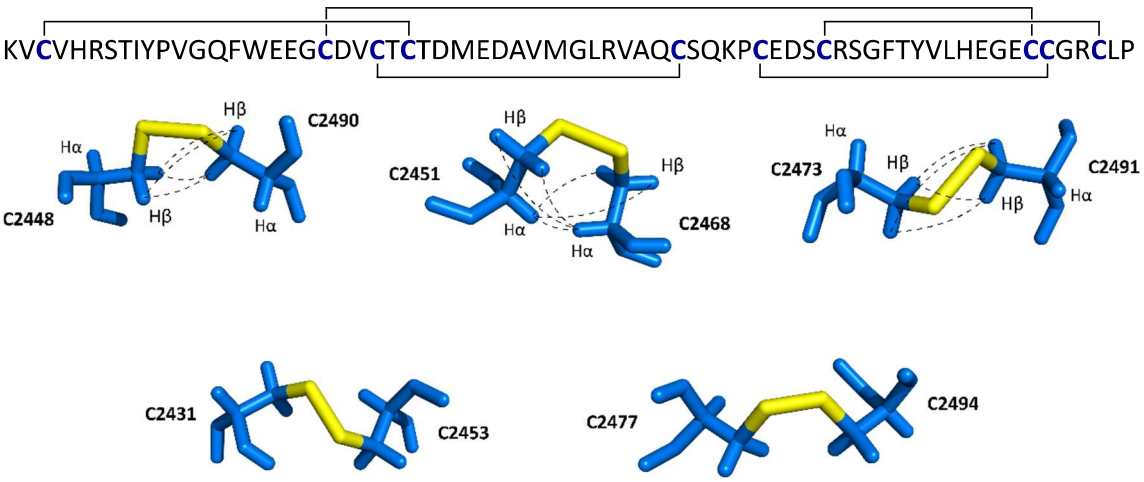


Figure 3.13: Arrangement of the VWF C3 disulphide bonds Indicated above is the disulphide bond arrangement in relation to the full VWF C3 sequence. Below are the five disulphide bonds formed in the final structure calculation. Dotted lines indicate inter-cysteine NOEs between $^{1}H_{\alpha}$ and $^{1}H_{\beta}$ nuclei. The two disulphide bonds between C2431-C2453 and C2477-C2494 do not contain any inter-cysteine NOEs, however, display NOEs to neighbouring residues.

Whilst a NOE assignment between two $^{1}H_{\beta}$ nuclei is ideal to determine a disulphide bond, as was seen between C2448-C2490 (Figure 3.9), overlapped peaks often make such a method impossible. When assigning a NOE between two alike nuclei, such as the $^{1}H_{\beta}$ of cysteine, they will inevitably have similar chemical shifts. For example, the chemical shifts of the $^{1}H_{\beta}$ nuclei of C2477 are 2.613 and 3.158 ppm, whilst the chemical shifts of the $^{1}H_{\beta}$ nuclei of C2494 are 2.752 and 2.925 ppm. With the signal from the diagonal peak inevitably being far stronger, as the $^{1}H_{\beta}$ nucleus corresponding to the cross-peak is separated by two sulphur atoms, the ability to distinguish between these peaks is extremely challenging. Overlap of the diagonal and cross-peaks becomes a more prominent issue when needing to focus on cross-peaks with a low intensity signal when zooming into the baseline noise. This difference in signal intensity between the diagonal and cross-peak can be clearly observed in Figure 3.9, where the signal for the C2490$^{1}H_{\beta b}$ cross-peak is far weaker, however, in this case is well separated, allowing a reliable NOE assignment between these nuclei.

Even though a NOE between the $^{1}H_{\beta}$ and $^{1}H_{\alpha}$ nuclei of two disulphide-bonded cysteines would tend to display even weaker cross-peaks than between $^{1}H_{\beta}$ nuclei, these cross-peaks are less likely to be overlapped. This was the case regarding a NOE between C2451$^{1}H_{\alpha}$ and C2468$^{1}H_{\beta}$ (Figure 3.8). These residues were expected to be disulphide bonded from the β-sheet

assignment, where they are embedded in the three-stranded β-sheet. This disulphide bond likely acts in stabilising this β-sheet structure. In a case where NOEs between cysteines are not possible, a NOE assignment between one cysteine and the nuclei of the neighbouring residue of its disulphide-bonded cysteine can provide an insight into a disulphide bond arrangement. This was observed between C2477 $^{1}H_{\beta a/b}$ and L2495 $^{1}H_{N}$ (Figure 3.10) and C2453 $^{1}H_{\alpha}$ and V2430 $^{1}H_{\gamma 1/2}$ (Figure 3.12). This proved particularly useful in the case of C2453 which displays extremely low peak intensity levels. This was observed throughout all NMR spectra, from the ^{15}N-HSQC to triple resonance experiments where NOEs from the $^{1}H_{\beta}$ nuclei were unobservable in the ^{13}C-edited NOESY spectrum, rendering NOE assignment between $^{1}H_{\beta}$ nuclei impossible. The fact that NOEs assigned by the method of using neighbouring residues are from two disulphides that would seemingly be located at opposite ends of the domain increases the confidence of these assignments. However, as is the case with each of these disulphide bond assignments, these were checked as the tertiary structural assignment progressed that they agreed with the results of other NOE assignments.

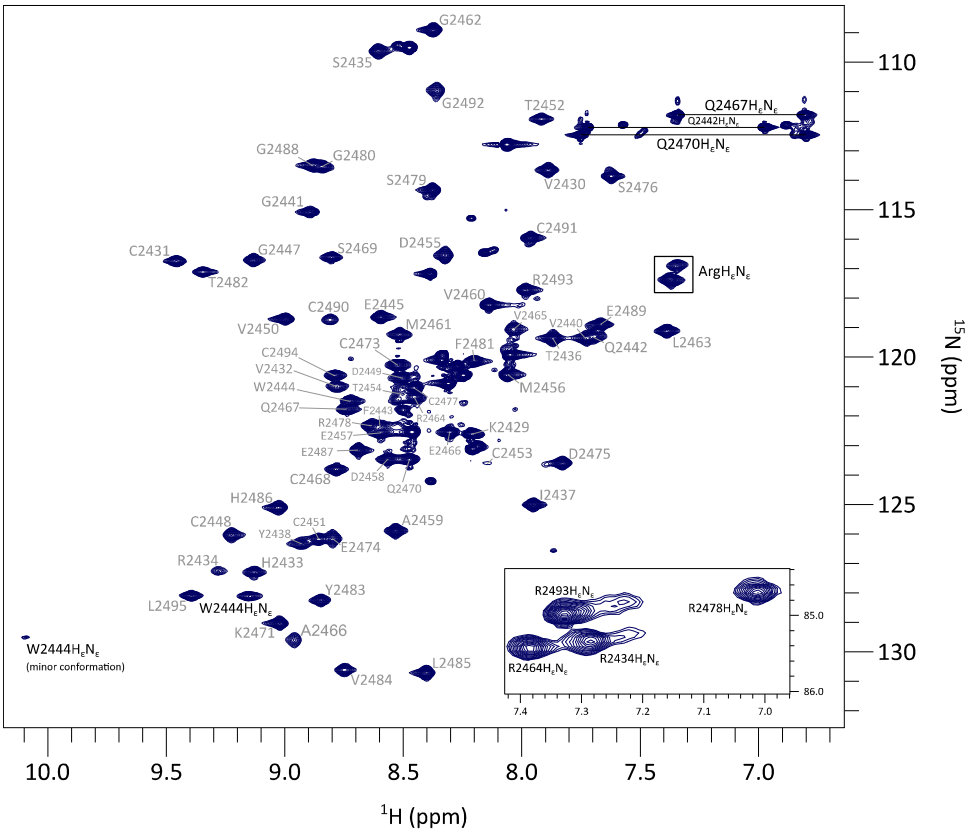


Figure 3.14: Assignment of the side chain amide groups of glutamine, tryptophan and arginine The VWF C3 ^{15}N-HSQC including the assignment of side chain ^{1}H-^{15}N groups with black labels showing the assignment of the aromatic $^{1}H_{\varepsilon}{}^{15}N_{\varepsilon}$ of W2444 and aliphatic $^{1}H_{\varepsilon}{}^{15}N_{\varepsilon}$ of arginine and glutamine residues. Side chain $^{1}H^{15}N$ groups of tryptophan and glutamine were assigned by a combination of the ^{13}C- and ^{15}N-edited NOESY, identifying intra-residue NOEs to determine the assignment. Labelled in grey are the backbone amide groups previously assigned. Labelled are the folded peaks corresponding to the $^{1}H_{\varepsilon}{}^{15}N_{\varepsilon}$ peaks of arginine residues and lower right is the assignment of these arginine nuclei. Assignment of the arginine $^{1}H_{\varepsilon}{}^{15}N_{\varepsilon}$ was performed by a side chain TOCSY in 20 mM Tris-HCl, 50 mM NaCl, pH 6.0.

Intra-residue NOE assignments were used to determine the $^{1}H_{\varepsilon}{}^{15}N_{\varepsilon}$ chemical shifts of glutamine and tryptophan (Figure 3.14). These contain NOEs to the aliphatic side chain nuclei of their respective residues and, in the case of W2444, to its neighbouring ^{1}H-^{13}C aromatic nuclei which were subsequently assigned. A second conformation of the aromatic $^{1}H_{\varepsilon}{}^{15}N_{\varepsilon}$ of W2444 was also assigned at a ^{1}H chemical shift of ~10.1 ppm. This chemical shift indicates an exposed tryptophan residue whereas the peak at ~9.1 ppm shows the residue in a buried conformation[415]. Comparing the intensity of the two peaks suggests that while W2444 exists in two conformations, its dominant conformation is buried, likely forming hydrophobic/hydrophilic interactions. The chemical shifts of the $^{1}H_{\varepsilon}{}^{15}N_{\varepsilon}$ nuclei of the four arginine residues were also assigned by a colleague, Diana Catana. Due to the spectral width of the ^{15}N-HSQC, these peaks are folded from a ^{15}N chemical shift of ~85 ppm. In order to assign these, two types of TOCSY experiments were performed, backbone and side chain, and by aligning the two spectra and the previously assigned peaks of the $^{13}C_{\alpha}$ to $^{13}C_{\delta}$ nuclei, each $^{1}H_{\varepsilon}{}^{15}N_{\varepsilon}$ peak was assigned.

Assignment of the aromatic side chain ^{13}CH and $^{13}CH_2$ groups of histidine, phenylalanine, tryptophan and tyrosine, aromatic amide group of tryptophan and side chain amide group of glutamine (no asparagine residues in VWF C3) were required for an accurate structure calculation. The aromatic side chain assignment is particularly important as these often provide stabilising interactions and, therefore, present strong NOEs to neighbouring residues, essential when determining the protein's tertiary structure. These were assigned using the ^{13}C-edited NOESY spectrum for the ^{13}CH and $^{13}CH_2$ groups (assignment of aromatic nuclei of Y2483 shown in Figure 3.15) and ^{15}N-edited NOESY for the amide groups. A second ^{13}C-edited NOESY experiment was carried out (indicated in Table 3.3) which focused on the ^{13}C chemical shift of the aromatic ^{13}C nuclei, in this case from 111.8 to 144.0 ppm rather than 9.3 to 87.0 ppm for the aliphatic nuclei.

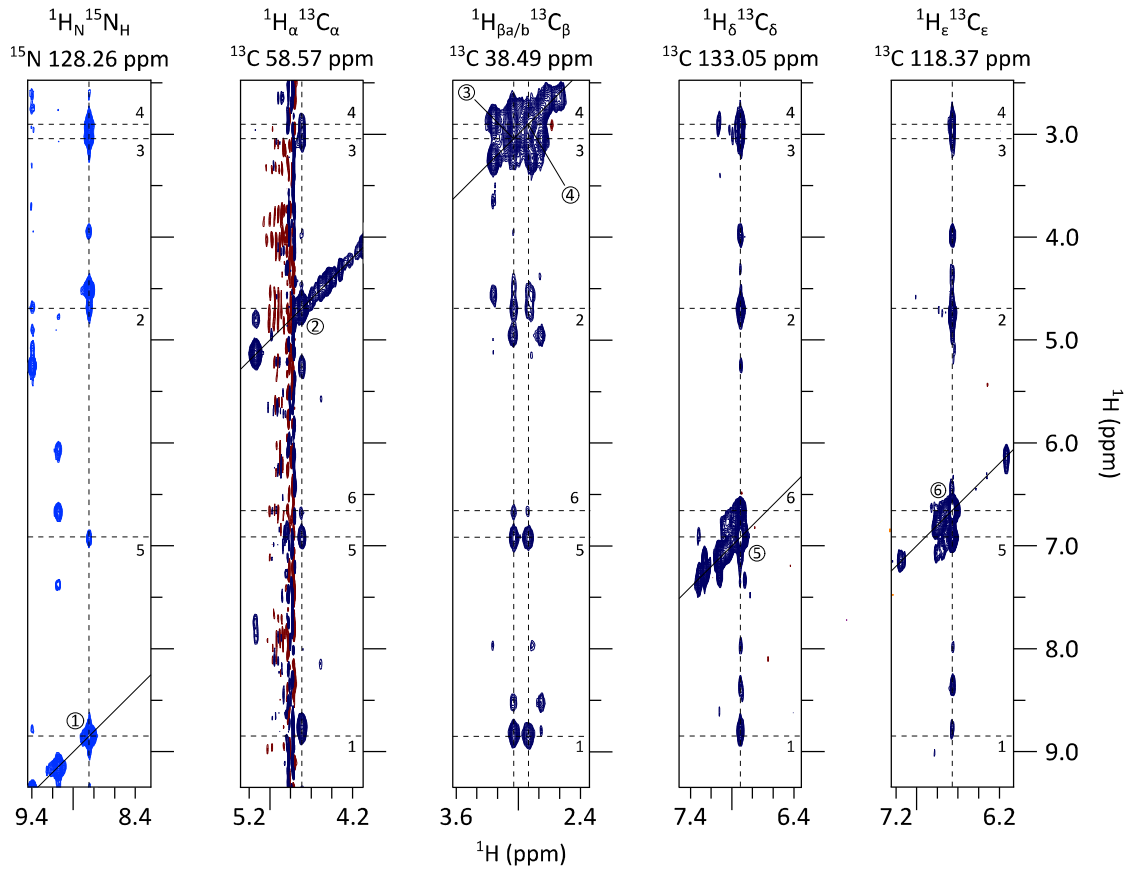


Figure 3.15: Assignment of the Y2483 aromatic nuclei Displayed is an example of the assignment of aromatic nuclei using the ^{15}N-edited (left strip) and ^{13}C-edited (other four strips) NOESY spectra. Throughout the spectra, blue corresponds to positive peaks and red to negative peaks. The diagonal peak in each strip corresponds to the nuclei labelled above its respective strip. The ^{15}N/^{13}C chemical shift at the top of each strip corresponds to the chemical shift in the z-dimension. Each diagonal peak is labelled with a circled number with the number to the right of each cross-peak corresponding to the diagonal peak it forms a NOE with. The two right-hand panels show the assignment of the $^{1}H_{\delta}{}^{13}C_{\delta}$ and $^{1}H_{\varepsilon}{}^{13}C_{\varepsilon}$ using the alignment of their cross-peaks with other intra-residue nuclei.

All aromatic ^{13}CH and $^{13}CH_2$ groups were determined using the NOESY spectra, including the aromatic $^{1}H_{\delta}{}^{13}C_{\delta}$ and $^{1}H_{\varepsilon}{}^{13}C_{\varepsilon}$ nuclei of Y2483 (Figure 3.15). The Y2483 aromatic assignment also shows the difference in signal strength when two nuclei are further apart. The $^{1}H_{N}$ nucleus cross-peaks (8.85 ppm) become noticeably weaker from the $^{1}H_{\alpha}{}^{13}C_{\alpha}$ strip to the $^{1}H_{\varepsilon}{}^{13}C_{\varepsilon}$ strip. This helps to distinguish between the $^{1}H_{\delta}$ and $^{1}H_{\varepsilon}$ nuclei, with the $^{1}H_{\delta}$ expected to show a stronger cross-peak to $^{1}H_{N}$ than $^{1}H_{\varepsilon}$. This assignment method was supplemented with two two-dimensional ^{1}H-^{13}C spectra, the HbCbCgCdHd and HbCbCgCdCeHe. These plot the chemical shift of $^{13}C_{\beta}$, a previously assigned chemical shift, against that of the intra-residue $^{1}H_{\delta}$ and $^{1}H_{\varepsilon}$, respectively. This was used to confirm the assignment and confidently distinguish between the $^{1}H_{\delta}{}^{13}C_{\delta}$ and $^{1}H_{\varepsilon}{}^{13}C_{\varepsilon}$. In the case of tyrosine, distinguishing between $^{13}C_{\delta}$ and $^{13}C_{\varepsilon}$ is relatively straightforward with the reference chemical shift of $^{13}C_{\delta}$ being ~133 ppm and that of $^{13}C_{\varepsilon}$ being ~118 ppm. However, those of phenylalanine are far more similar with $^{13}C_{\delta}$ having a

reference chemical shift of ~131.5 ppm and $^{13}C_{\varepsilon}$ being ~131 ppm. Correlating the $^{13}C_{\beta}$ nucleus with the $^{1}H_{\delta}$ and $^{1}H_{\varepsilon}$ proves particularly useful in cases such as phenylalanine. Furthermore, in the case of two aromatic residues being in close proximity, assignment using NOESY spectra would prove challenging, with difficulty arising when distinguishing between intra- and inter-residue NOEs. The HbCbCgCdHd and HbCbCgCdCeHe spectra prevent a mis-assignment in this case with these only providing intra-residue correlations. The assignment of the $^{1}H_{\varepsilon}{}^{13}C_{\varepsilon}$ nuclei of M2456 and M2461 was also assigned using the ^{13}C-edited NOESY spectrum, identifying intra-residue NOEs.

This resulted in a final assignment of 97.4% (710 of 729 nuclei). This high percentage contributes to an accurate structural determination. This data does not include those ^{13}C nuclei that are not ^{1}H-bound, such as the $^{13}C_{\gamma}$ nucleus of the aromatic residues, likewise ^{1}H nuclei that are not ^{13}C- or ^{15}N-bound, such as $^{1}H_{\eta}$ of tyrosine or $^{1}H_{\gamma1}$ of threonine (O-bound) nor histidine or lysine side chain $^{1}H^{15}N$ groups. These would all require further experiments to determine their chemical shifts and an accurate structure calculation can be performed with the chemical shift data obtained thus far. Whilst the disulphide bonds provide an initial insight into the tertiary structure, inter-residue NOEs were an essential element to finalise the overall domain folding.

NOEs between H2433, Y2438 and W2444 (Figure 3.16) indicate an interaction between the first two β-sheet structures. Both H2433 and Y2438 reside in the first β-sheet (Figure 3.6, Panel A) with W2444 situated in the first strand of the second β-sheet structure (Figure 3.6, Panel B). This suggests a possible stabilising interaction between these two β-sheets, however, this will become clearer when assessing the calculated structure. These interactions are between aromatic nuclei, highlighting the importance of assigning these nuclei for an accurate structure calculation. The minor, exposed conformation of W2444 where $^{15}N_{\varepsilon}$ has a chemical shift of ~10.1 ppm does not display any NOEs to neighbouring residues, further indicating a lack of stabilising interactions by W2444 in this conformation.

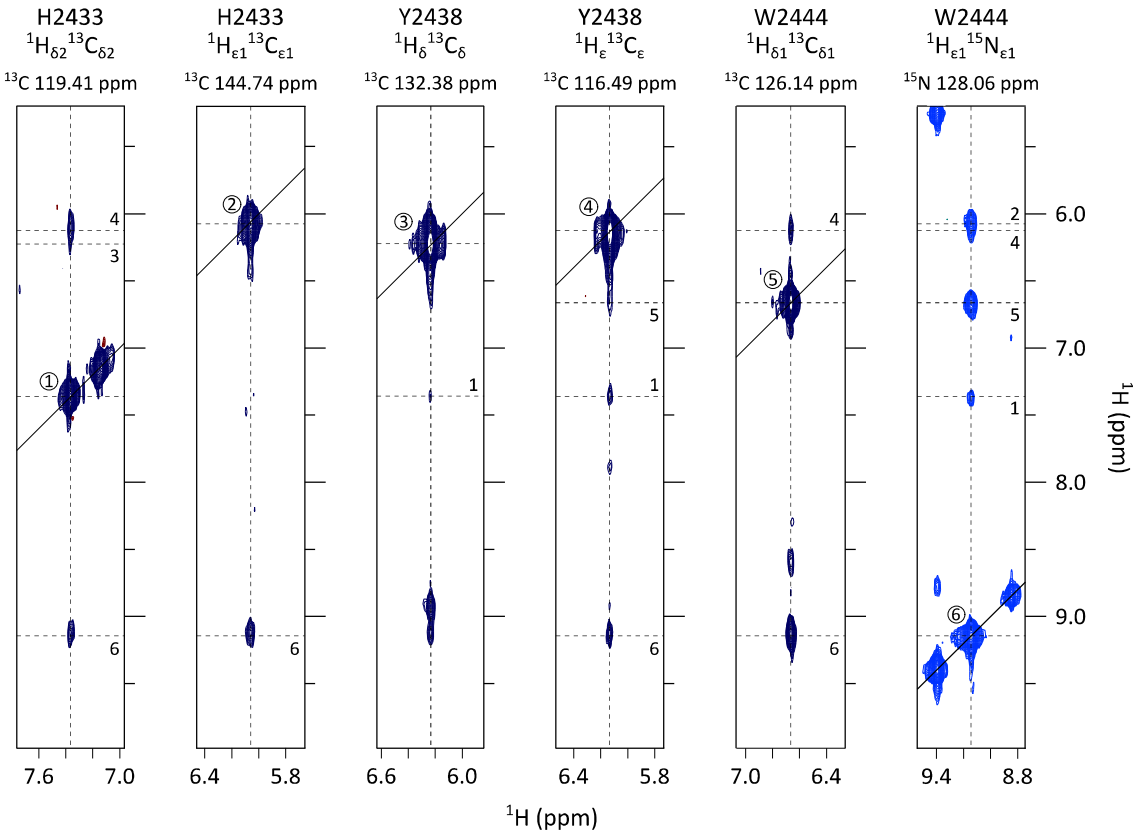


Figure 3.16: Tertiary structure NOE assignment between residues H2433, Y2438 and W2444 A NOE assignment determining the tertiary structure between residues of the first (H2433 and Y2438) and second (W2444) β-sheet structures. The first five strips are using the ^{13}C-edited NOESY and the last from the ^{15}N-edited NOESY. Labelled nuclei above each peak corresponds to the diagonal peak in that strip, with the ^{13}C chemical shift above each strip corresponding to the ^{13}C chemical shift in the z-dimension. Diagonal peaks are labelled with circled numbers and each cross-peak labelled with the number of the diagonal peak that it corresponds to.

Tertiary structure NOE assignments can also be used to justify the disulphide bond assignments made previously, particularly for those where NOEs between two cysteine residues was not possible. The NOEs assigned in Figure 3.17 reinforce the assignment of C2431-C2453 with R2464 and A2466 in close proximity to C2453. This is known through the β-sheet assignment in Figure 3.6 (Panel B) where C2453 immediately proceeds the second strand of the β-sheet and R2464 and A2466 immediately precede the third strand of the β-sheet.

Likewise in Figure 3.18**,** a NOE between R2478 and F2481 justifies the disulphide bond assignment between C2477 and C2494 which was previously assigned by a NOE between C2477 and L2495. Immediately preceding the first strand of the β-sheet in Figure 3.6 (Panel C) is F2481 and in close proximity to C2494. With F2481 forming a NOE with the neighbouring residue to C2477 (R2478), it further suggests that these two residues are disulphide bonded.

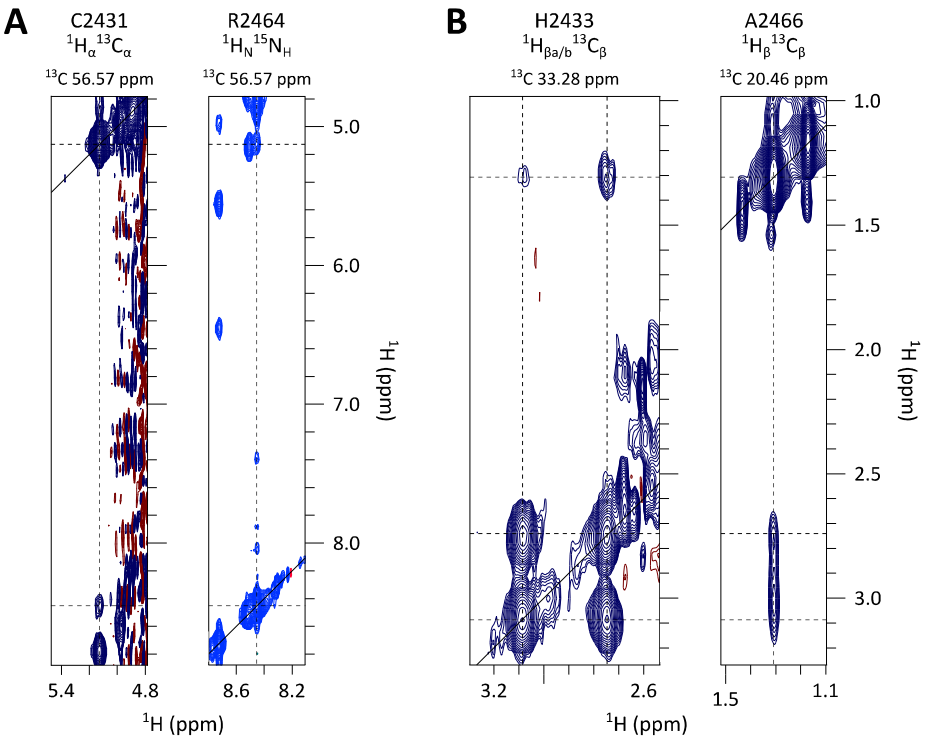


Figure 3.17: Tertiary structure NOE assignment A NOE assignment between C2431 and R2464, Panel A, and H2433 and A2466, Panel B. The second strip in Panel A displays a ^{15}N-edited NOESY whilst the other three strips are the ^{13}C-edited NOESY. Nuclei labelled above correspond to the diagonal peak in its corresponding strip. The ^{13}C chemical shift labelled above each strip is the chemical shift in the z-dimension. Cross-peaks in each strip form a NOE with its aligned diagonal peak.

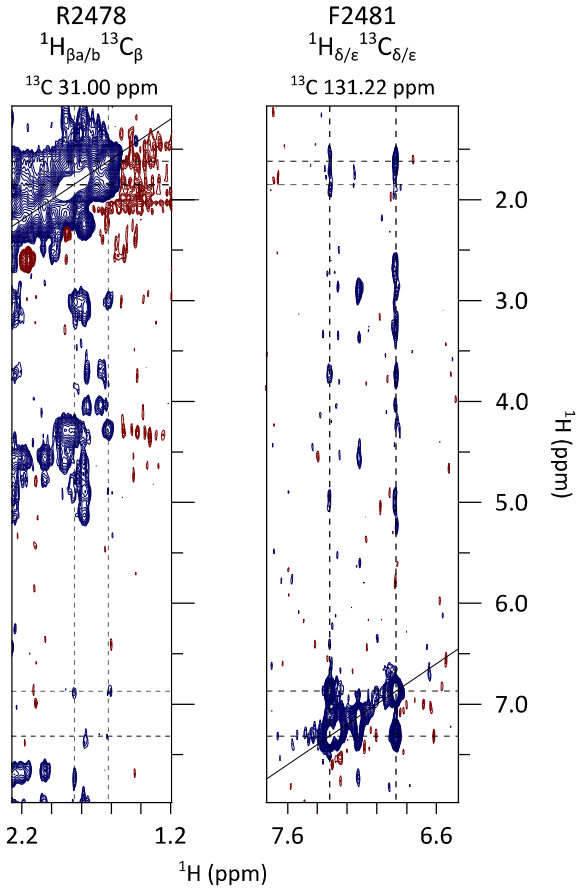


Figure 3.18: NOE assignment between R2478 and F2481 The assignment of a NOE between the $^{1}H_{\beta}$ nuclei of R2478 and the $^{1}H_{\delta}$ and $^{1}H_{\varepsilon}$ nuclei of F2481 using the ^{13}C-edited NOESY. Blue peaks are positive whilst red peaks are negative. Diagonal peaks correspond to the nuclei labelled above each strip with cross-peaks correlating with the diagonal peaks they form a NOE with.

The NOEs described in this chapter provide sufficient information, along with the chemical shift data, for an accurate structure calculation. A total of 2642 NOE restraints were assigned, 1224 of which were inter-residue NOEs. It was ensured that any inter-residue NOEs, particularly long-distance, were supported with neighbouring NOEs, preventing an incorrect fold from the structure calculation. Removal of any NOEs was reviewed throughout the structure calculation process to optimise the dataset. The inter-residue NOEs include 723 sequential NOEs, 89 short-range (two to three residue separation), 69 medium-range (four to five residue separation) and 343 long-range (more than five residue separation) (Figure 3.19). The NOE assignment is dominated by sequential assignments with long, medium and short-range NOEs concentrated in the β-sheet regions. Outside of these secondary structure regions, NOEs are prevalent around the disulphide bonds, however, longer-range NOEs are limited. This would suggest, as expected, that the C3 domain is not a globular structure with a hydrophobic core as in the A domains as a core would present a far greater number of long-range NOEs within that structure. A definitive conclusion, however, would only be possible following the structural calculation. Also notable in the NOE assignment was a lack of NOEs between the linker impurity mentioned in Chapter 2.8 to the C3 structure. This further indicates that the linker impurity is not incorporated into the VWF C3 structure, rather residing as an unstructured tail.

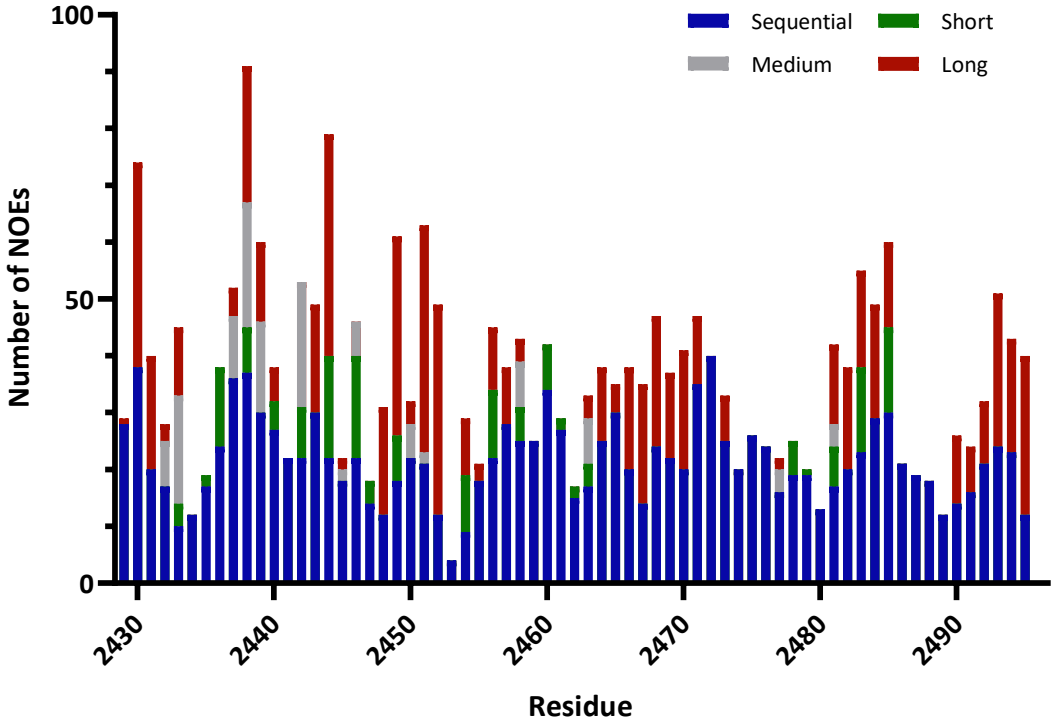


Figure 3.19: Assigned NOEs for VWF C3 residues Long (more than five residue separation), medium (three or four residue separation), short-range (two or three residue separation) and sequential (neighbouring residues) NOEs assigned for each VWF C3 residue. Sequential NOEs are shown in blue with short, medium and long-range NOEs in green, grey and red, respectively.

3.4. Residual Dipolar Couplings

Whilst an accurate structural assignment can be completed using the information obtained thus far, refining the orientations of bond vectors within an overall molecular frame increases the precision and accuracy of a protein structure. This is particularly significant in the case of asymmetric structures and multidomain proteins which are dominated by α-helices or β-sheets, as has been learnt is the case in VWF C3[416]. This refinement includes the measurement of residual dipolar couplings (RDCs), with an RDC between two nuclei reporting on the internuclei vector orientations. In this case we looked at the coupling between the ^{1}H and ^{15}N nuclei of the backbone amide group in order to improve our structure. When a protein in solution is exposed to a magnetic field, the protein weakly aligns relative to that field. This results in a dipole-dipole interaction which yields a measurable dipolar coupling (Figure 3.20). The magnitude of the coupling depends on the angle between the internuclei vector and the magnetic field. The coupling is also affected by the distance between the nuclei, however, in this case where the ^{1}H-^{15}N nuclei are covalently bound, this distance is fixed[417]. The alignment in this case, however, is minute unless an alignment media is added.

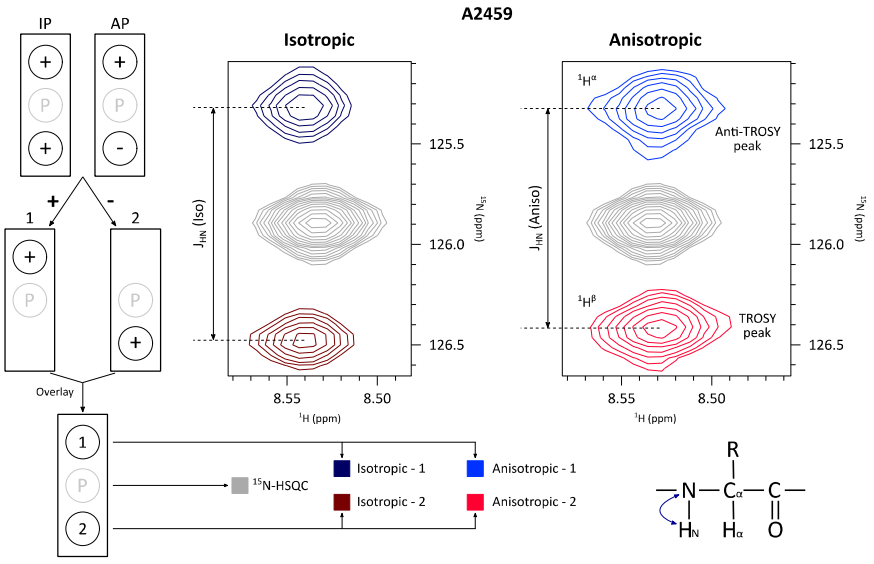


Figure 3.20: Residual dipolar coupling measurements using an in-phase/anti-phase (IPAP) experiment
The left panel shows the processing of the IPAP ^{15}N-HSQC experiment in order to measure the $^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$ couplings (J) (bottom right). This experiment produces doublets in the ^{15}N dimension, splitting a ^{15}N-HSQC peak both up and downshift as the ^{1}H-^{15}N coupling is allowed to evolve in the ^{15}N chemical shift evolution period. When the splitting is in-phase (IP) it results in two positive peaks whilst the anti-phase (AP) splits into a positive and negative peak. By adding the two spectra, the positive and negative upshift peaks cancel each other out, only displaying the downshift Anti-TROSY peak (1). By subtracting the two spectra, only the upshift TROSY peak is observed (2) and the splitting between the two measured by overlaying these resulting spectra. This method prevents crowding of a spectrum due to the resulting doublets. Shown in the two NMR spectrum strips are the ^{1}H-^{15}N splitting of A2459 with the left strip being isotropic and the right strip anisotropic. The coupling (J) is indicated as the measurement between the two peaks, with an RDC measured by the difference between the isotropic and anisotropic coupling.

The addition of an alignment media creates an anisotropic matrix. The interaction between the protein and the media can produce weak alignment of the protein[417] and the alignment media used in this case was bacteriophage *Pf*1. This bacteriophage consists of a circular genome within an α-helical coat protein structure. The repeating network of carbonyl groups in the coat protein causes the bacteriophage to have a large anisotropic magnetic susceptibility with a solution of these particles imparting alignment of the protein to which the bacteriophage *Pf*1 is added to[418]. Whilst in isotropic conditions dipolar couplings average to zero with each possible orientation having an equal probability, upon addition of an alignment media and subsequent anisotropic conditions, a preference of a certain orientation will be observed. In the case of the isotropic coupling, this remains relatively consistent across all amide groups, varying in VWF C3 between 91 Hz and 96 Hz with the average coupling being 93.6 Hz. In anisotropic conditions, however, this coupling varies between 78 Hz and 103 Hz. The anisotropic coupling was measured in triplicate and the mean taken for each amide group. By measuring the difference between the coupling of an amide group in isotropic and anisotropic conditions, an RDC can be measured for each backbone amide group. Measurement of RDCs was not possible for proline residues due to their lack of an amide ^{1}H nucleus, F2443 and T2454 due to peak overlap and C2453 due to a lack of signal.

3.5. High Resolution Structure of VWF C3

A structure of VWF C3 was determined using CYANA 3.0[419] (**c**ombined assignment and d**y**namics **a**lgorithm for **N**MR **a**pplications) and the NMR restraints calculated thus far. These include the assignment of 2642 intra- and inter-residue NOEs, 114 phi and psi torsion angle restraints predicted from TALOS+, five disulphide bonds, 62 ^{1}H-^{15}N RDCs and 64 $^{13}C_{\alpha}$-$^{13}C_{\beta}$ RDCs. Phi and psi torsion angles refer to the backbone dihedral angles with phi corresponding to the angle formed between the backbone C_O-N_H-C_α-C_O nuclei and psi between the backbone N_H-C_α-C_O-N_H nuclei. These phi and psi torsion angle restraints were discounted for residues at hairpin loop regions. A CYANA structure calculation includes the calculation of the distance between nuclei for each observable NOE. This calculation also delivers a number of distance and backbone dihedral angle violations. In order to optimise the calculation, the distance restraints for each NOE were eased by adding the initial distance restraint to the maximum violation distance. The easing of these violations has been shown to be imperative as a small number of incorrect restraints can lead to a completely misfolded protein[420]. Following subsequent structure calculations, this process was repeated until distance violations were eliminated. To rectify backbone torsion angle violations, the calculated structure was analysed

using the Protein Structure Validation Software (PSVS) suite[421]. This provided a Ramachandran plot for each residue and where violations occurred, elucidating the extent of the violation and the change in angle restraint required. The process calculates 100 potential structures from the restraints provided, finally presenting the twenty lowest energy structures. In these results, we present these twenty lowest energy structures with the lowest energy structure presented in Figure 3.21.

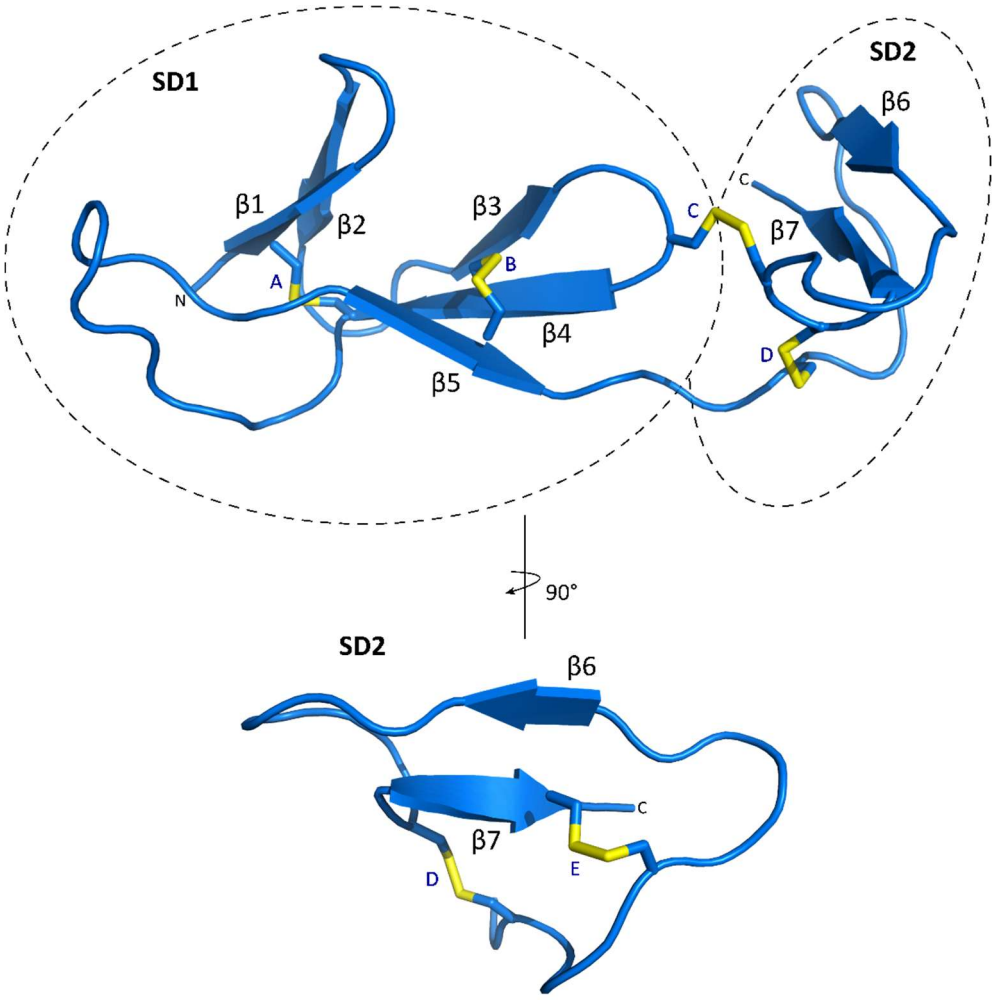


Figure 3.21: High resolution structure of the VWF C3 domain Displayed above is a ribbon diagram of the isolated VWF C3 domain, showing two individual subdomains, SD1 and SD2. Indicated are the seven β-sheet strands, β1-5 in SD1 and β6 and β7 in SD2. Also shown are the five disulphide bonds (A-E) with the S_γ nuclei of cysteine residues shown in yellow. Disulphide bonds are between C2431-C2453 (A), C2451-C2468 (B), C2448-C2490 (C), C2473-C2491 (D) and C2477-C2494 (E). Shown below is SD2 when rotated by 90°, providing a clear picture of its β-sheet (β6 and β7) and disulphide bond arrangement (D and E). The lowest energy structure is presented here.

The solution structure of the VWF C3 domain is presented in Figure 3.21. The domain is comprised of two subdomains, SD1 and SD2, with seven β-sheet strands, β1-β7. The first double-stranded (β1 and β2) and the triple-stranded β-sheet (β3-β5) are in SD1 with the second double stranded β-sheet (β6 and β7) in SD2. The angle between the N-terminal β-sheet

and remainder of SD1 is maintained by a π-stacking interaction between the aromatic side chains of Y2438 and W2444. This is also stabilised by the first of two disulphide bonds found in each SD1 with the C2431-C2453 (A) bond stabilising the interaction between the two β-sheets in SD1. The second disulphide bond in SD1, C2451-C2468 (B), strengthens the secondary structure interaction between strands β4 and β5. The two disulphide bonds in SD2 (D and E) connect the unstructured lower strand in SD2 to the β7 strand. VWF C3 also contains an inter-subdomain disulphide bond (C) connecting the β3-to-β4 loop to the β6-to-β7 loop. This arrangement indicates that the two subdomains are independently folded with a hinge residue at K2471 and a disulphide bond between C2448 and C2490 stabilising their orientation with respect to each other. With the lack of NOEs between the subdomains it appears that K2471 and the C2448-C2490 disulphide bond maintain the orientation of the two subdomains.

As was expected following analysis of similar VWFC domains and assignment of the NOESY spectra, VWF C3 lacks any hydrophobic core structure with all residues solvent exposed. Its stable structure, on the other hand, is dominated by its regions of secondary structure (47% of residues) and intervening loop regions. These loop regions vary from short hairpin loops connecting β-sheet strands, as in the β1-to-β2 loop, to long, protruding loops, such as that between strands β4 and β5. Although lacking any α-helix regions, TALOS+ predicted a small strand centred around V2460 (Figure 3.5, Panel C), immediately preceding strand β5. Although not recognised as an α-helix, this region does display a turn akin to α-helical properties. As is the case with a number of other VWF domains due to their high percentage of cysteine residues, the disulphide bonds are the primary stabilising force in the VWF C3 structure, as was predicted by TALOS+ where the most rigid regions of the domain are centred around the cysteine residues (Figure 3.7).

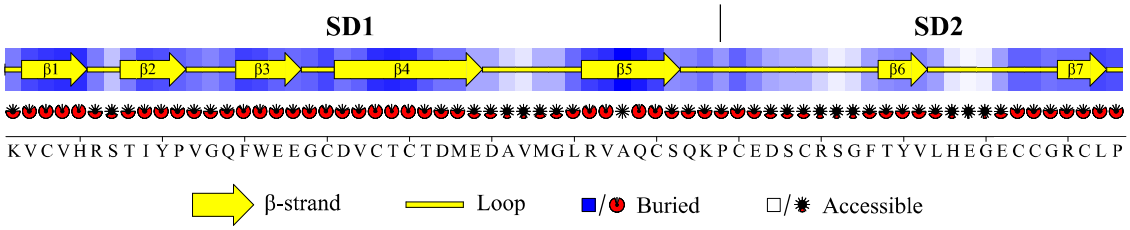


Figure 3.22: Accessibility of VWF C3 residues The displayed graphic the accessibility of the individual residues of VWF C3 based on the twenty lowest energy structures generated by Procheck-NMR. Indicated by yellow arrows are regions of β-sheet strands, marked β1-β7, and loop regions indicated by yellow lines. Blue and closed red circles signify buried residues whereas accessible residues are shown in white and black star segments. The Ooi number, the number of C_α atoms within a 14 Å radius of each amino acid, is used to determine their accessibility.

Despite lacking a hydrophobic core, Figure 3.22 indicates that VWF C3 contains a number of buried residues, particularly within the regions of secondary structure in SD1. Whilst SD2 contains buried regions around residues F2481-H2486 and S2490-P2496 which house strands β6 and β7, respectively, the remainder of the subdomain is far more accessible than that of SD1. Particular accessible regions seen throughout the domain are exposed loop regions, with the long β4-to-β5 loop, the loop preceding strand β6 and the loop connecting strand β6 to C2490 which forms an intersubdomain disulphide bond. Cysteine residues tend to appear as buried residues, highlighting their role in forming core stabilising disulphide bonds.

The RMSD was measured of the twenty lowest energy structures of VWF C3 to indicate how well the individual structures of the ensemble agree with each other. In order to improve the precision of the bundle, the RMSD was minimised, indicating minimal variation between the individual structures and increasing the reliability of the calculated structure[422]. The mean backbone RMSD of the VWF C3 protein is 0.525 Å, with a range of 0.257 to 0.784 Å over the twenty lowest energy structures. However, when analysing the separate subdomains, SD1 has an average RMSD of 0.170 Å whilst SD2 has an average RMSD of 0.400 Å. This suggests that SD2 is more dynamic than SD1, as can be seen when observing the twenty lowest energy structures and the greater variation in SD2 conformation than SD1 (Figure 3.23).

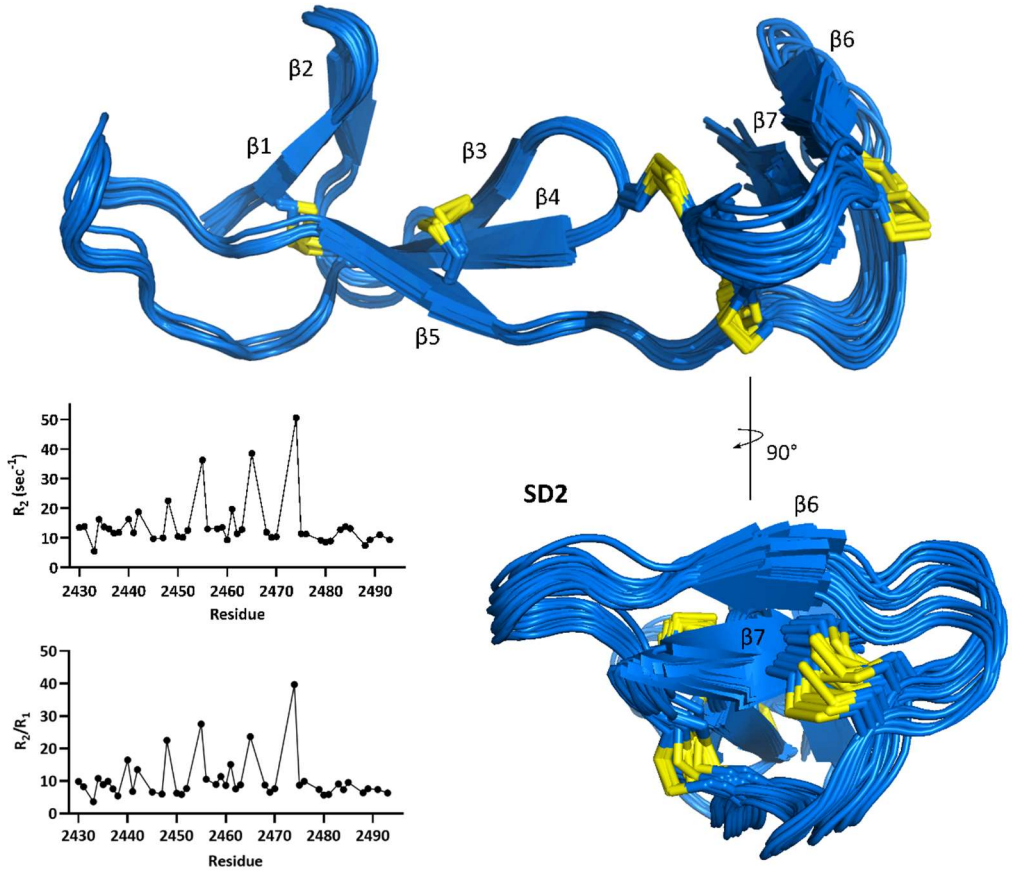


Figure 3.23: Twenty lowest energy structures of VWF C3 Displayed is a ribbon diagram of the twenty lowest energy structures of VWF C3, with the lower panel showing the structure rotated through 90° showing SD2. Both structures are in the same orientation as that in Figure 3.21. Disulphide bonds are displayed with the C_α, C_β and S_γ nuclei of cysteine residues shown as sticks with S_γ marked in yellow. The higher degree of variation in SD2 conformation than SD1 indicates greater flexibility in SD2, in agreement with the RMSD values of the two subdomains. Also displayed is the R_2 relaxation data across the C3 domain derived from ^{15}N relaxation data along with experimental R_2/R_1 values.

Conclusions drawn from RMSD values were supplemented with ^{15}N relaxation data to confirm regions of rigid and flexible structures. Similar to that found in C4 relaxation data[125], the β1-to-β2 loop around H2433 exhibits fast dynamics in the nanosecond to picosecond time scale, shown by a decreasing R_2/R_1 value. A significantly higher R_2/R_1 value is observed for E2474, indicating a flexible hinge residue with E2474 undergoing slow conformational exchange in the microsecond to millisecond time scale. This residue, E2474, also contributes to the perpendicular orientation of SD2 in relation to SD1, with its acidic side chain forming a salt bridge with the basic side chain of R2478 (Figure 3.24). Finally, the consistently lower R_2/R_1 values in SD2 in comparison to SD1 indicate higher global mobility in SD2, in agreement with RMSD data, exhibiting a greater number of orientations.

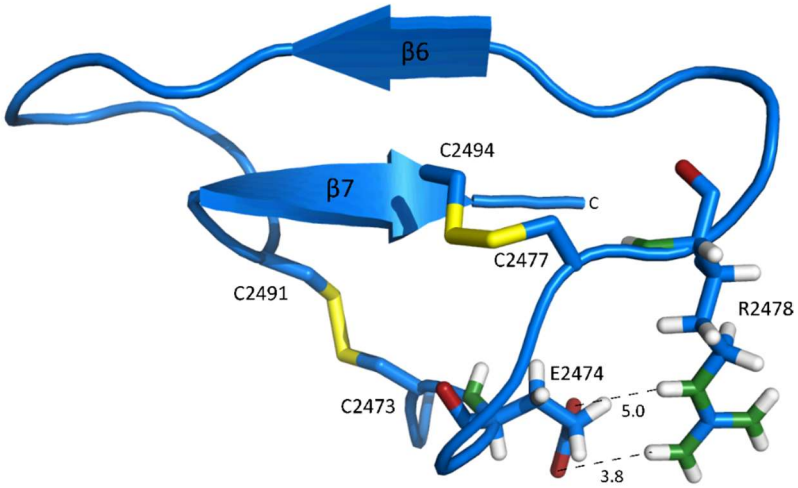


Figure 3.24: Salt bridge between E2474 and R2478 The interaction between E2474 and R2478 in SD2 is displayed with a salt bridge forming between the CO_δ of E2474 and the NH_η of R2478. This interaction contributes to the change in orientation of SD2 with respect to SD1, causing a 90° turn in SD2. Dotted lines indicate the interaction between the CO_δ nuclei of E2474 and NH_ε and NH_η which are 5.0 Å and 3.8 Å apart, respectively. To ensure an accurate representation of these distance measurements, they were calculated as an average across the twenty lowest energy structures.

Assessing the twenty protein structures using a Ramachandran plot determines the quality of the calculated structures by checking its geometry and stereochemistry. A Ramachandran plot of a protein's phi-psi torsion angles indicates the percentage of residues in the most favoured, allowed and disallowed regions[423]. The Ramachandran plot of the calculated VWF C3 structure presents 68.5% in most favoured regions, 30.7% in additionally allowed regions and 0.8% in generously allowed regions, with no residues in the disallowed regions (Appendix 7.2). With VWF C3 being a small disulphide-rich protein lacking a hydrophobic core and having a less regular secondary structure, the proportion of residues in the most favoured regions is lower than would be seen in larger proteins with a prominent secondary structure[424].

3.6. Structural comparison to VWFC domains

Analysis of the sequence relationships between VWF C3 and other VWFC domains indicates that the VWFC domain of CCN3 GF is that with most similarity to the VWF C3 domain. The NMR structures of the two domains show similarly low RMSD values, with C3 exhibiting an average RMSD of 0.525 Å and 0.265 Å in the CCN3 GF VWFC domain. Whilst an overlay of the two full-length domains does not show the two structures matching well, this is due to differences in the SD2 orientation of the two domains. Overlay of solely the SD1 domains, however, shows strong structural similarities (Figure 3.25). Both contain two β-sheet regions, one double-stranded and one triple-stranded with almost identical β-sheet lengths and loop

regions from strand β1 to β4. The orientation of the N-terminal β-sheet to the remainder of SD1 is also identical with both domains containing a β1 to β4 disulphide bond (Figure 3.26) and a π-stacking interaction between the aromatic side chains of a tyrosine and tryptophan residue. The β4-to-β5 loop is where the domains' SD1 differ, with VWF C3 containing a far longer loop than the CCN3 GF VWFC domain. The functional significance of the long β4-to-β5 loop in the VWF C3 domain as yet remains unclear, however, with the 26 residue C2 to C3 domain linker, there is a possibility the longer β4-to-β5 loop in C3 stabilises the interaction with the VWF C2 domain.

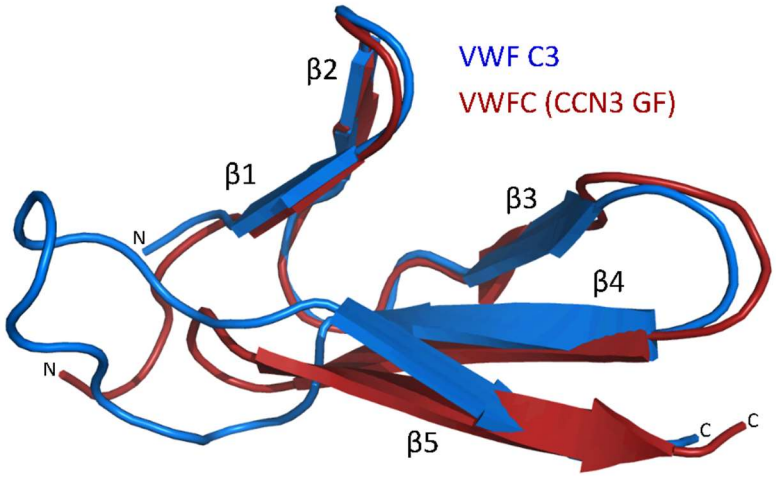


Figure 3.25: Overlay of the SD1 of VWF C3 and the VWFC domain of CCN3 GF Displayed is SD1 of the VWF C3 domain (blue), overlaid with the CCN3 GF VWFC domain (red). Both contain five β-sheet strands with similar secondary structure arrangements, the one significant difference between the two being the long β4-to-β5 loop.

With the CCN3 GF VWFC domain also being a small, disulphide-rich domain lacking a hydrophobic core, the arrangement of its five disulphide bonds is similarly vital to forming a stable tertiary structure. As can be seen in Figure 3.26, both VWF C3 and the CCN3 GF VWFC domain have similar disulphide bond arrangements. Both have two disulphides in SD1, with one connecting the two β-sheets between strands β1 and β4 and one stabilising the triple-stranded β-sheet between strands β4 and β5. In both cases, the two subdomains are connected by a hinge residue and a disulphide between the β3-to-β4 loop and the β6-to-β7 loop, stabilising the association of the two subdomains. Variation, however, occurs in SD2, firstly in the disulphide bond arrangement. Whilst one disulphide bond of SD2 is identical in both domains (Figure 3.26, Bond D), the final disulphide bond (Bond E) varies between the two. In the C3 domain, this bond is between strand β7 and the unstructured lower strand, similar to Bond D, however, in the CCN3 GF VWFC domain this bond connects strand β7 to the loop immediately preceding strand β6. The lack of a disulphide bond between the two β-sheet

strands in SD2 of the VWF C3 domain may affect the stability of this secondary structure. Whilst this bond in the CCN3 GF VWFC domain is not directly between strands β6 and β7, its position may still play a role in stabilising the SD2 secondary structure.

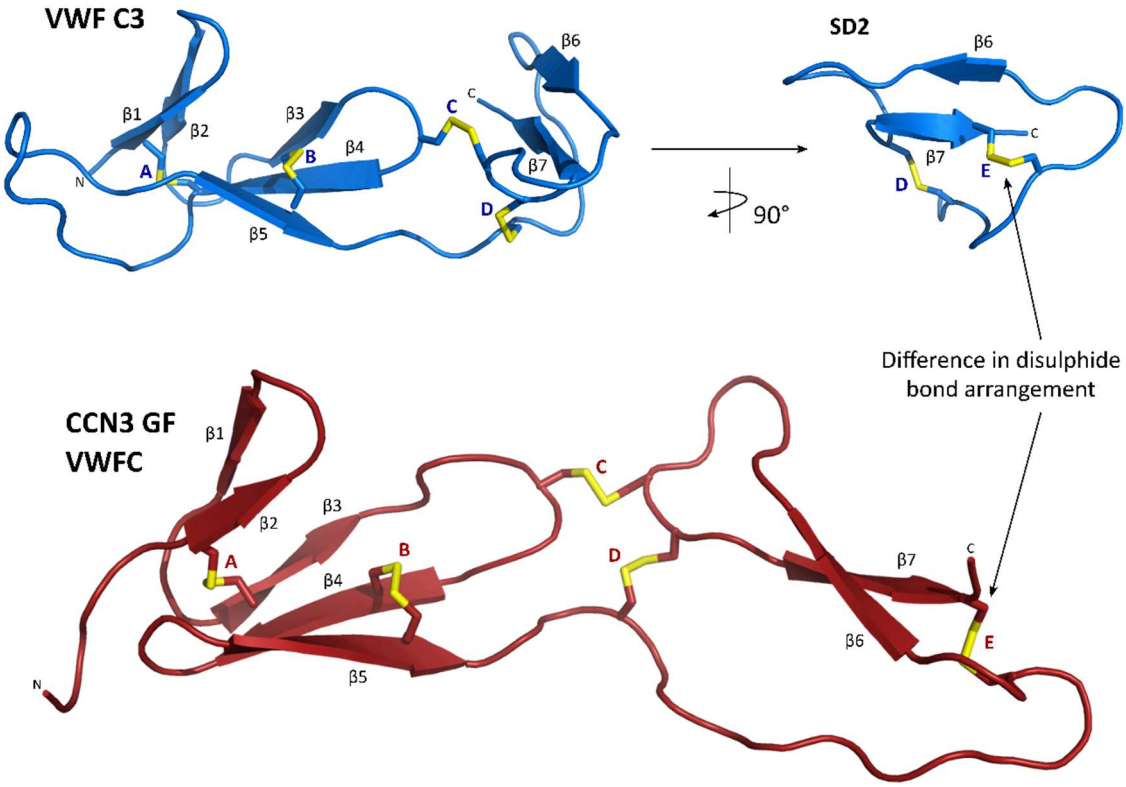


Figure 3.26: Structural comparison between the full-length VWF C3 domain and CCN3 GF VWFC domain Displayed in the above panel in blue is the full-length VWF C3 domain with an end-on view of SD2 shown in the upper right segment. Below is the VWFC domain of CCN3 GF, shown in red. In both cases, sticks are shown for the C_α, C_β and S_γ of cysteine residues, with the S_γ marked in yellow. Disulphide bonds are labelled *A* to *E*, with the labels between the two proteins matching with their equivalent disulphide bonds in the other protein.

The final two noticeable differences between the two domains are also presented in SD2. Firstly, the size of SD2 is noticeably larger in the VWFC domain of CCN3 GF. The VWF C3 domain SD2 (P2472 to P2496) is 25 residues in length whilst the VWFC domain of CCN3 GF is 31 residues in length, allowing the CCN3 GF VWFC domain to be more elongated than that of VWF C3. The final significant difference between the two domains is the orientation of SD2 in relation to SD1. As shown in Figure 3.24, SD2 of C3 lies perpendicular SD1 rather than longitudinal caused by the interaction of the acidic and basic amino acid side chains of E2474 and R2478, respectively, leading to a salt bridge[425]. A turn in the protein chain is the result of this interaction, causing this perpendicular orientation. Whilst the CCN3 GF VWFC domain contains a similar acidic residue, D141, in an equivalent position to E2474 in VWF C3, it does not contain a basic residue to form this salt bridge interaction and rotate SD2.

3.7. Structural comparison to the VWF C4 domain

Whilst the VWF C4 domain contains fewer sequence similarities to the VWF C3 domain than the VWFC domain of CCN3 GF, it is insightful to compare the structures of the two neighbouring domains. The C4 domain displays a similar secondary and tertiary structure arrangement to C3 (Figure 3.27). Whilst its disulphide bond arrangement is comparably similar to C3, it only contains one disulphide bond in SD1, between strands β1 and β4, lacking the second disulphide bond between strands β4 and β5. It does, however, contain a third disulphide bond in SD2 with its unique C4 bridge between strands β6 and β7, similar to the final disulphide bond of the CCN3 GF VWFC domain (Figure 3.26, Bond E).

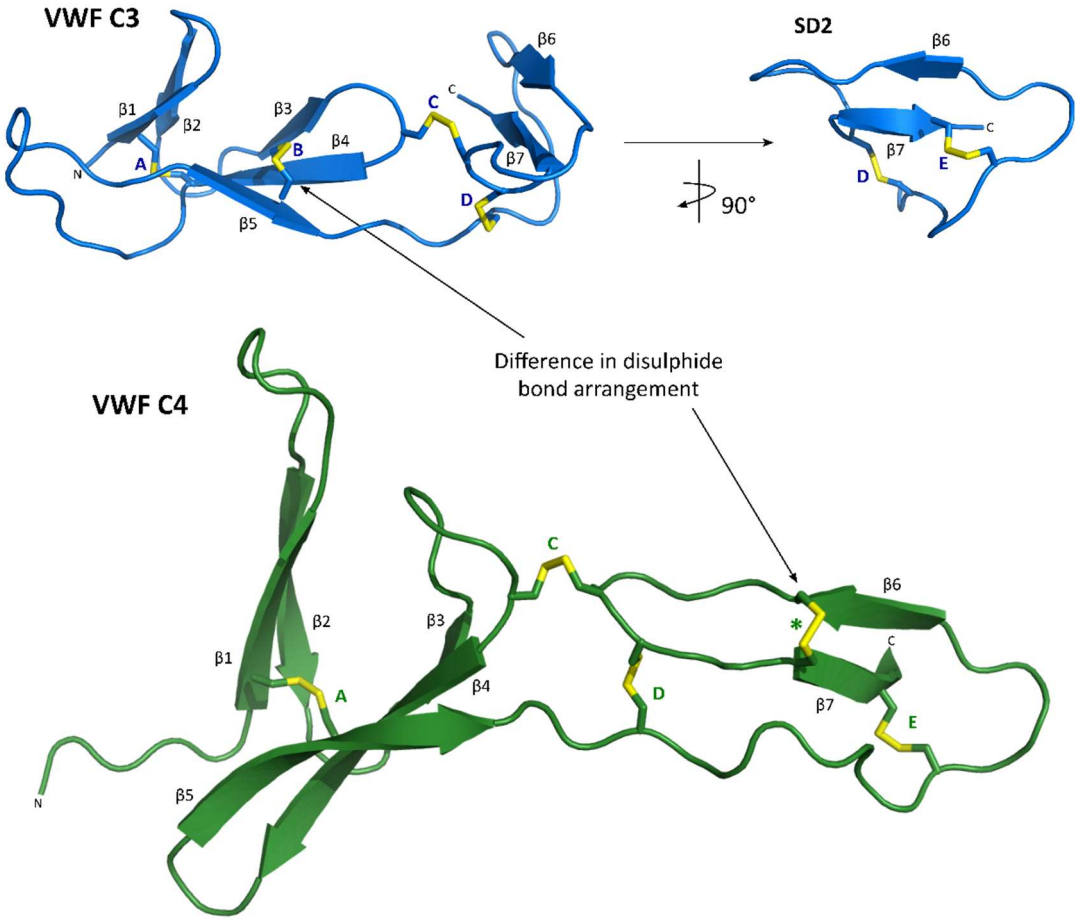


Figure 3.27: Structural comparison between the full-length VWF C3 domain and VWF C4 domain
Displayed in the above panel in blue is the full-length VWF C3 domain with an end-on view of SD2 shown in the upper right segment. Below is the VWF C4 domain[125], shown in green. In both cases, sticks are shown for the C_α, C_β and S_γ of cysteine residues, with the S_γ marked in yellow. Matching disulphide bonds between the two domains are labelled *A*, *C*, *D* and *E*, with C4 lacking disulphide bond *B* which is shown in C3. The C4 domain also contains a third disulphide bond in SD2 marked with a *.

The orientation of SD2 in the VWF C4 domain is akin to that in the CCN3 GF VWFC domain, residing longitudinally in relation to SD1. Similar to the VWFC domain of CCN3 GF, VWF C4

contains an acidic residue, E2553, in the equivalent position to E2474 in VWF C3, however, lacks the basic arginine residue that is found in VWF C3. The lack of a salt bridge in this position prevents the turning of the domain and the change in orientation seen in VWF C3.

When comparing the structure of SD1 of VWF C3 and C4, it is immediately apparent why VWF C4 is termed a VWFC-like domain. Whilst it contains a number of structural hallmarks that strongly associates it with the VWFC domain family, particularly its secondary structure arrangement, an overlay of SD1 of VWF C3 and C4 presents some clear differences (Figure 3.28). In general, the size of the β-sheets are far larger, with strands β1, β2, β4 and β5 all stretching to approximately double the length of those in VWF C3. It also contains a much larger β1-to-β2 loop which contains an RGD motif that binds to platelet $\alpha_{IIb}\beta_3$.

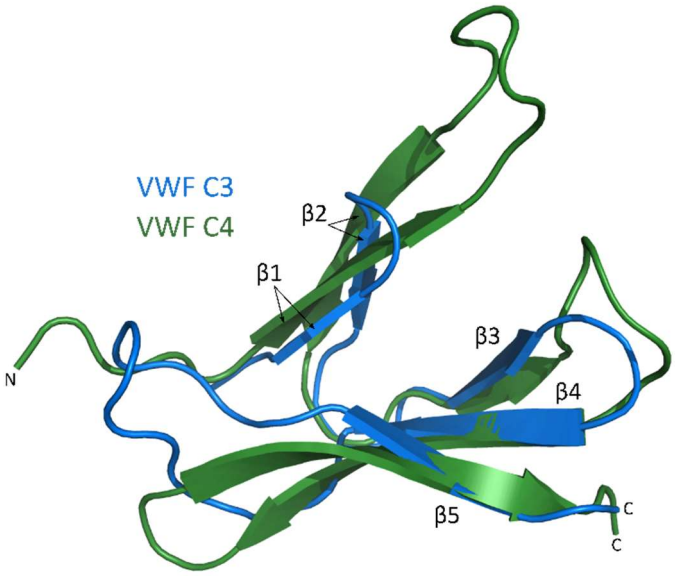


Figure 3.28: Overlay of the SD1 of the VWF C3 and C4 domains Displayed is a structural comparison of the first subdomain of the VWF C3 and C4 domains. VWF C3 is shown in blue whilst the C4 domain is shown in green.

The C4 β1-to-β2 loop is both longer than the hairpin loop seen in C3 (Figure 3.28) as well as far more flexible, taking on a wider range of conformations in the twenty lowest energy structures of C4 (Figure 3.29). On the other hand, this loop in VWF C3 takes a much more consistent conformation throughout its twenty lowest energy structures. This difference is understandable with the flexibility in C4 surrounding the RGD motif likely being essential for efficient binding to $\alpha_{IIb}\beta_3$. This is also seen in ^{15}N relaxation data where, although H2433 of the C3 domain shows fast dynamics, it is not to the same extent as that seen in the much larger β1-to-β2 loop of the C4 domain. The comparison of the twenty lowest energy structures also displays the difference in SD2 flexibility in VWF C3 and C4. This is apparent with the much

wider variety of conformations seen in SD2 of the C4 domain. Whilst SD2 of the C3 domain appears to be less flexible than that of C4, it does not display the same conformation consistency as seen in SD1. This is also seen in the C4 domain where, although the β1-to-β2 loop is highly flexible, the β3-β4-β5 secondary structure has a highly consistent conformation between the twenty lowest energy structures. The experimental ^{15}N relaxation data of both C3 and C4 also suggests greater flexibility in SD2 than SD1, with SD2 consistently showing lower R2/R1 values indicating higher global mobility. This feature is a hallmark of the VWFC domains, consisting of a uniformly stable SD1 and SD2 displaying greater flexibility. Such a feature may contribute to the intra-domain flexibility and, subsequently, multimer flexibility under shear stress.

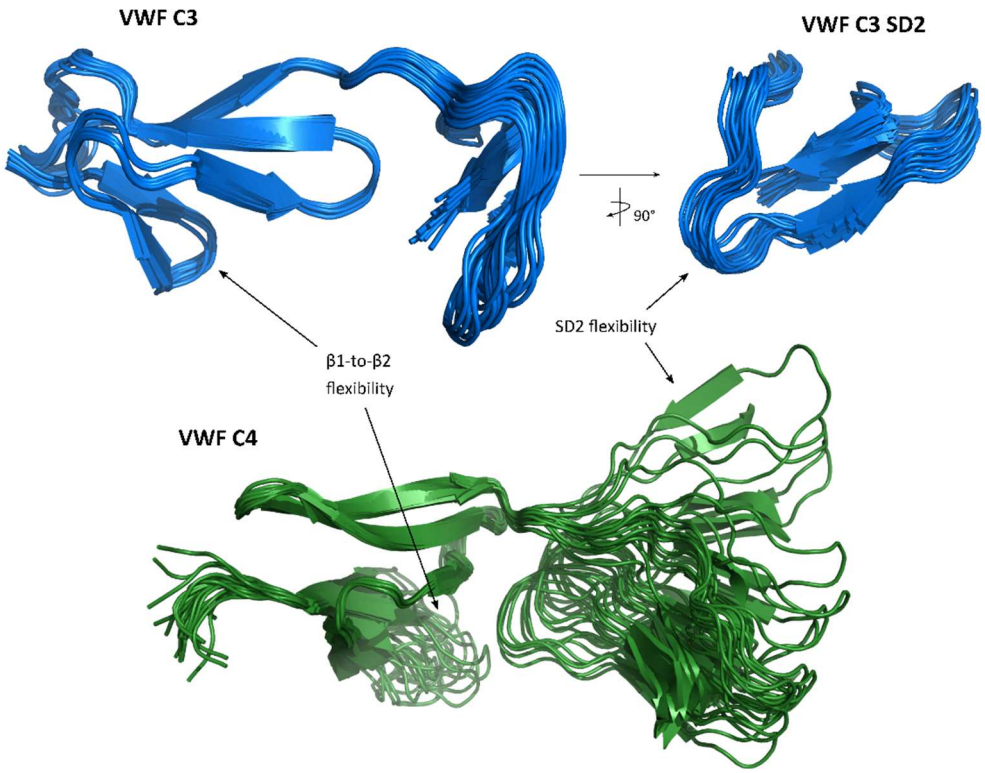


Figure 3.29: Comparison of the twenty lowest energy structures of VWF C3 and C4 Displayed above is a comparison of the VWF C3 and C4 structures, displaying their differences in flexibility throughout the domains. Indicated are those regions with differing flexibility, with the β1-to-β2 loop and SD2. VWF C3 is shown in blue and VWF C4 in green. The upper right section shows SD2 of the C3 domain rotated through 90°.

3.8. Interaction between the VWF C3 and C4 domains

Whilst the two neighbouring domains display similar structural arrangements, the interactions between these two domains remains to be determined. Their interaction, or lack of, would have a profound effect on, firstly, inter-domain flexibility and, secondly, flexibility of the VWF

multimer as a whole. A key role of the C domains is to add flexibility to the VWF multimer with the remainder of the protein being either highly disulphide bonded (D domains) or globular with a hydrophobic core (A domains). The inter-domain junctions and intra-domain flexibility is what provides flexibility to the VWF multimer in order for it to effectively carry out primary haemostasis.

In order to investigate the interactions between the two domains, the ideal method of expressing the two domains in complex was not possible. A plasmid, of the same form of the VWF C3 plasmid, was expressed with a TrxA tag, a His_6 tag for NiNTA purification, an EK cleavage site, immediately followed by the VWF C3-C4 sequence. Due to the fact that the C3-C4 complex contains ten disulphide bonds, five in each domain, the expression was extremely inefficient. Multiple attempts were carried out and a minimal concentration of uncleaved TrxA-His_6-VWFC4 was expressed. As was feared, the number of disulphide bonds was too high for such an expression strategy. Instead, the C3 and C4 domains were expressed separately.

A plasmid (Figure 3.30) encoding for the VWF C4 domain was designed and expressed.

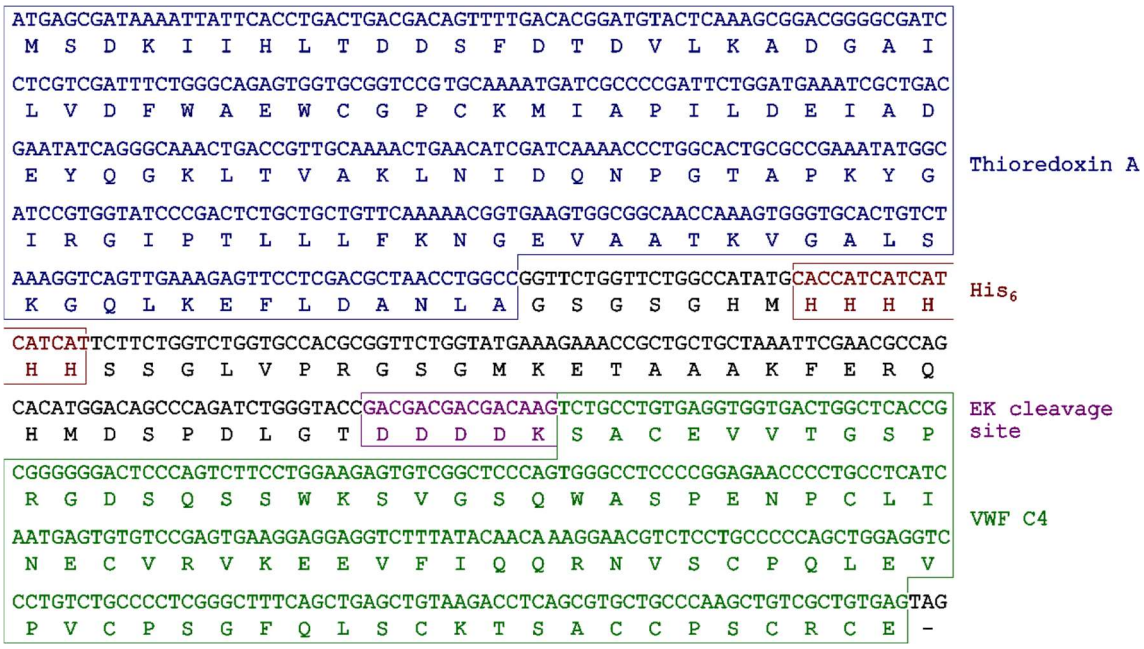


Figure 3.30: TrxA-His_6-VWFC4 construct sequence The nucleotide sequence and subsequent translation product to express the VWF C4 domain. The 109 amino acid TrxA tag (blue) is separated from a His_6 tag (red) by 7 residues, with the His_6 tag followed by a 31 residue linker to an EK cleavage site (purple). The cleavage site is immediately followed by the VWF C4 protein sequence (green) which is subsequently followed by a termination codon.

The protein was expressed and purified in the same manner as with the C3 plasmid, expressed in M9 minimal media with $^{15}NH_4Cl$ and $^{12}C_6$-$^{1}H_{12}$-glucose and the uncleaved TrxA-His_6-VWFC4

purified using Ni-NTA affinity chromatography. Whilst the uncleaved did not express concentrations as high as that from the TrxA-His_6-VWFC3 plasmid, it was far superior to that expressed in the C3-C4 domain complex. The purified protein was cleaved by EK and purified, firstly using Ni-NTA affinity chromatography to separate the cleaved C4 protein from the TrxA-His_6 tag and any uncleaved protein, and secondly using size exclusion chromatography to obtain a pure sample of VWF C4. The SDS-PAGE analysis of the purification and the size exclusion chromatogram are shown in Figure 3.31.

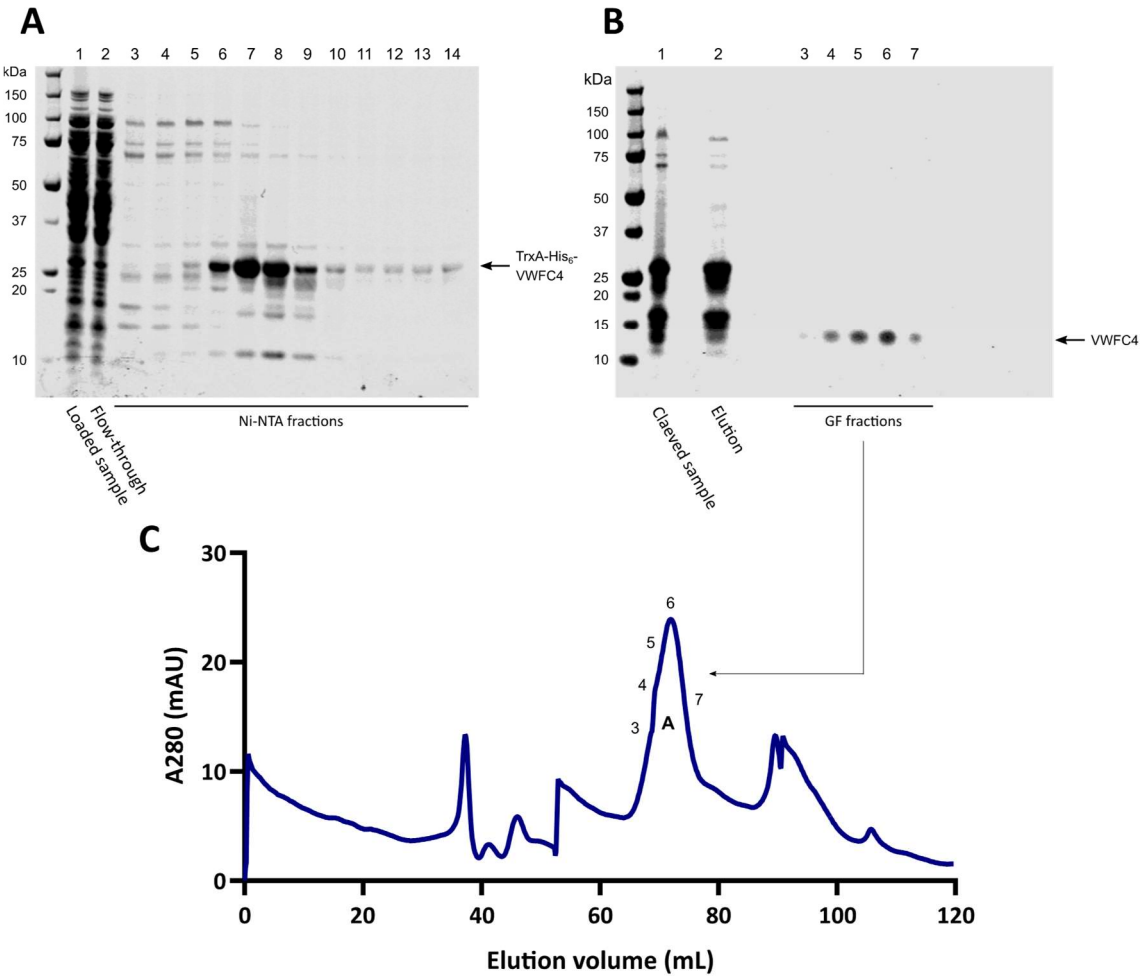


Figure 3.31: Purification of ^{15}N-labelled VWF C4 The Ni-NTA affinity chromatography purification of the uncleaved TrxA-His_6-VWFC4 is shown in Panel A, analysing the elution fractions in reducing SDS-PAGE. Lane 1 shows the contents extracted from the cell cytoplasm, with Lane 2 showing the unbound protein that flowed through the column. Fractions eluted from the column are shown in Lanes 3-14. Panel B analyses the elution fractions from the size exclusion chromatography (Panel C), following a reverse Ni-NTA. Lane 1 shows the cleaved protein that was loaded onto the column, with the bound protein that was subsequently eluted from the column shown in Lane 2. The unbound protein in the flow-through was separated by size exclusion chromatography. Lanes 3-7 correspond to the numbers seen on the chromatogram in Panel C, with the large band at 13 kDa corresponding to cleaved VWF C4.

Whilst the size of the protein eluted appeared too large for the 8.8 kDa C4 domain, it was likely that the same issue regarding cleavage had occurred as with the VWF C3 domain. The eluted

protein, however, was analysed by a ^{15}N-HSQC and compared to the VWF C4 domain published by Xu et al[125]. The peak assignment was successfully transferred onto the C4 domain presented here, increasing confidence that a folded structure of VWF C4 had been expressed. As the aim of this was to identify any obvious interaction between VWF C3 and C4, a protein with the addition of a N-terminal unstructured chain was not deemed an issue.

To identify any obvious interactions, a ^{15}N-labelled sample of VWF C3 and C4 were added together at equal concentrations and a ^{15}N-HSQC performed. The spectrum was compared to the respective C3 and C4 ^{15}N-HSQC spectra (Figure 3.32), with any peak shift providing an insight into interactions between the two domains.

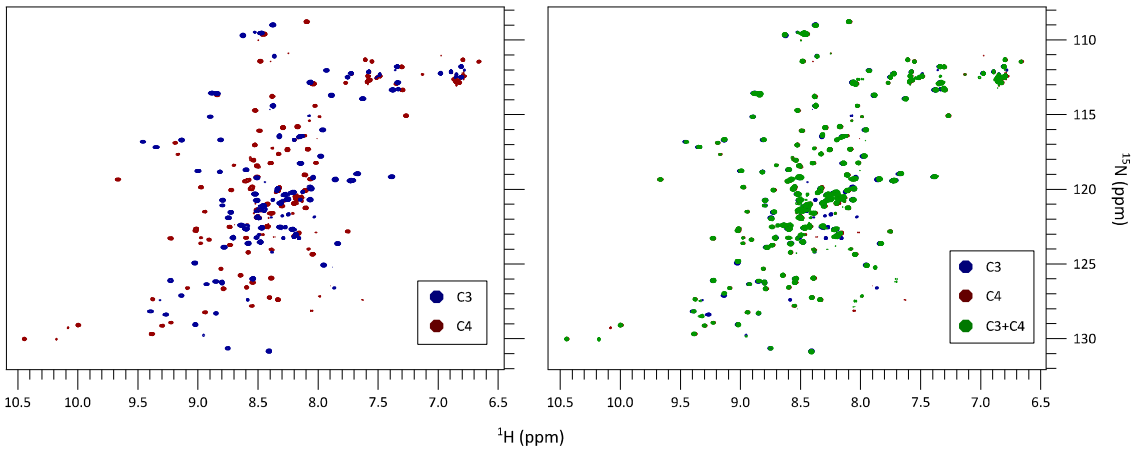


Figure 3.32: ^{15}N-HSQC overlap of the VWF C3 and C4 domains Displayed in the left-hand window are the ^{15}N-HSQC spectra of VWF C3 (blue) and C4 (red). In the right-hand window is an overlay when the C3 and C4 samples were added together (green).

There was no obvious peak shift from the ^{15}N-HSQC of the C3 and C4 domains when the two samples were added together. Minor peak shift was observed in the lower left quadrant of the spectrum corresponding to the backbone H2433 $^{1}H_{N}{}^{15}N_{H}$ and the W2444 $^{1}H_{\varepsilon}{}^{15}N_{\varepsilon}$. The significance of this is unclear, particularly how these two residues are in close contact between sheets β1 and β3. However, the fact that the pH of the two samples (pH 6.4) is so close to the pK_a of histidine (~6.0), the likelihood of a peak shift from H2433 is very high. This data provides an initial understanding of the lack of interaction of these two domains and whether there is any obvious stabilising interactions between the two domains. It is not, however, conclusive in determining their interaction with this requiring further investigation. It must also be considered the differences in interaction between these domains, and others, when VWF is in circulation in its globular form versus under shear stress and elongated. Whilst in the elongated form, any stabilising interactions would not be expected due to the importance of inter-domain flexibility, however, stabilising ionic interactions and salt bridges may occur in the

globular form of VWF. There is seemingly a lack of literature specifically regarding the interactions of the C domains whether in circulation or under shear stress, therefore, the perturbation of one domain on the other, or lack thereof, would require further investigation. Evidence does, however, highlight the importance of the C domains in contributing to the great length and flexibility of VWF multimers with the C domains exhibiting a rope-like character[37, 191]. Any strong inter-domain interactions would adversely affect such flexibility which is essential in the physiological function of VWF.

In order for a reliable conclusion to be drawn, the domains would firstly need to be expressed in complex. If expressing the two domains in complex is too troublesome due to the high number of disulphide bonds, this may require expressing the two proteins in parallel and performing a ligation reaction to join the two domains. Such a method, however, would not be possible with the purified protein containing a large unstructured "tail" at its N-terminus.

3.9. Expression of the C3L plasmid

Due to the inaccurate cleavage observed when expressing the VWF C3 domain and the unstructured "tail" sequence that remained at the N-terminus of the C3 domain, a new plasmid was designed which contains a four amino acid (GSMA) linker between the EK cleavage site and the C3 protein sequence. The remainder of the plasmid was identical to that shown in Figure 2.16. These residues were chosen due to their use when solving the NMR structure of the VWF C4 domain[125]. This would provide greater accessibility for EK to access the cleavage site, resulting in an accurate cleavage and a far smaller N-terminal "tail" structure. The protein was expressed and purified using the same method as previously (Figure 3.33).

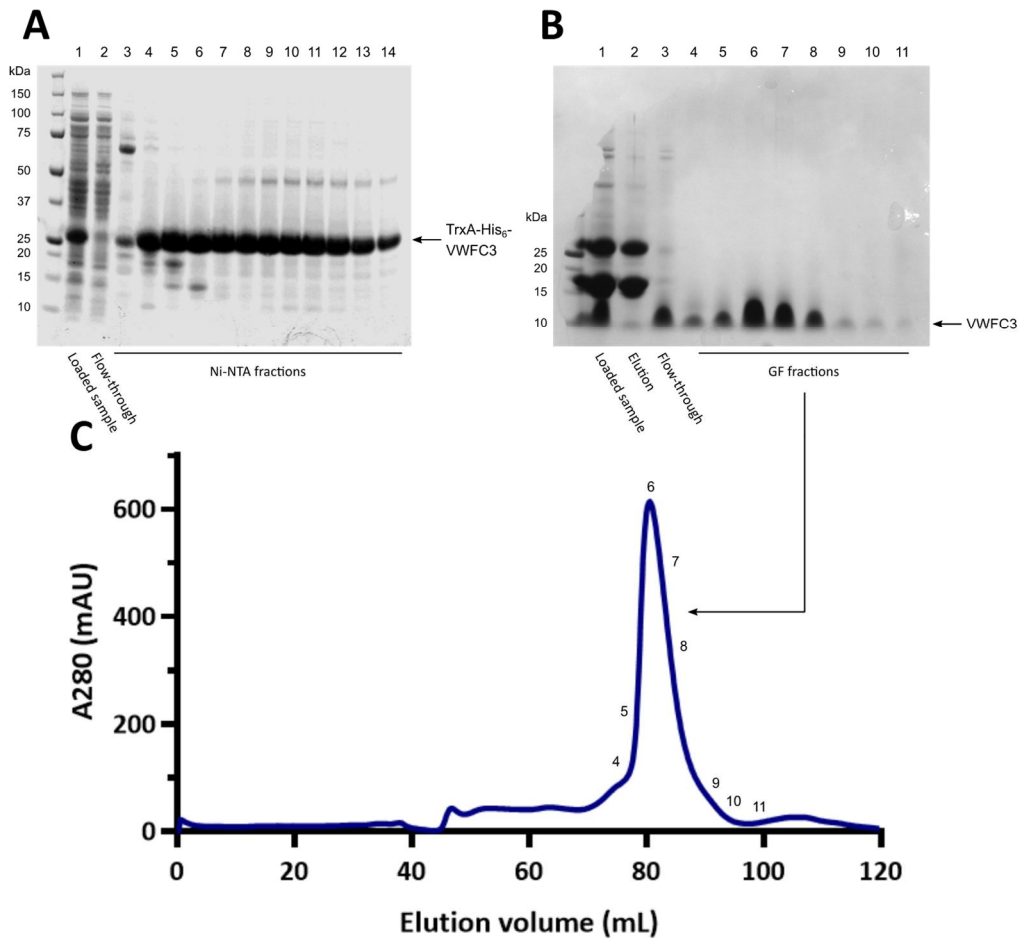


Figure 3.33: Purification of the C3L protein Shown in Panel A is the Ni-NTA affinity chromatography purification of the uncleaved TrxA-His_6-VWFC3 (C3L plasmid), analysing the elution fractions in reducing SDS-PAGE. Lane 1 shows the contents extracted from the cell cytoplasm, with Lane 2 showing the unbound protein that flowed through the column. Fractions eluted from the column are shown in Lanes 3-14. Panel B analyses the elution fractions from size exclusion chromatography (Panel C), following a reverse Ni-NTA. Lane 1 shows the cleaved protein that was loaded onto the Ni-NTA column, with the bound protein that was subsequently eluted from the column shown in Lane 2. The unbound protein in the flow-through, shown in Lane 3, was separated by size exclusion chromatography. Lanes 4-11 correspond to the numbers seen on the chromatogram in Panel C.

The purification results immediately suggested that EK cleavage had occurred at the intended site, firstly with the C3 protein peak centred at an elution volume of 80.5 mL compared with 74.5 mL when inaccurate cleavage had occurred, and secondly with the eluted protein displaying a size lower than 10 kDa. The expected size of the cleaved protein is ~8 kDa. Though difficult to determine its exact size by the SDS-PAGE analysis partly due to the high concentration of protein and spread of the band, the size is clearly less than 10 kDa. Furthermore, the high concentration of purified protein was encouraging for further studies such as a ligation with neighbouring VWF C domains in order to investigate their interaction.

In order to compare the folding of the C3 protein expressed from the C3L plasmid to the C3 structure calculated earlier, a ^{15}N-HSQC was performed (Figure 3.34) and backbone chemical shift changes were quantified. This would also provide an insight into whether the unstructured "tail" from the previous plasmid had had an effect on the calculated C3 structure. To differentiate between the expressed protein of the two plasmids, the C3 protein used for the original structure calculation will hereby be referred to as C3NL (indicating "No Linker" between the cleavage site and C3 protein) and that expressed from the new C3L plasmid will be referred to as C3L (denoting the "Linker" following the cleavage site). The chemical shift changes between C3NL and C3L ($\Delta\delta$) were determined by the equation in Figure 3.35.

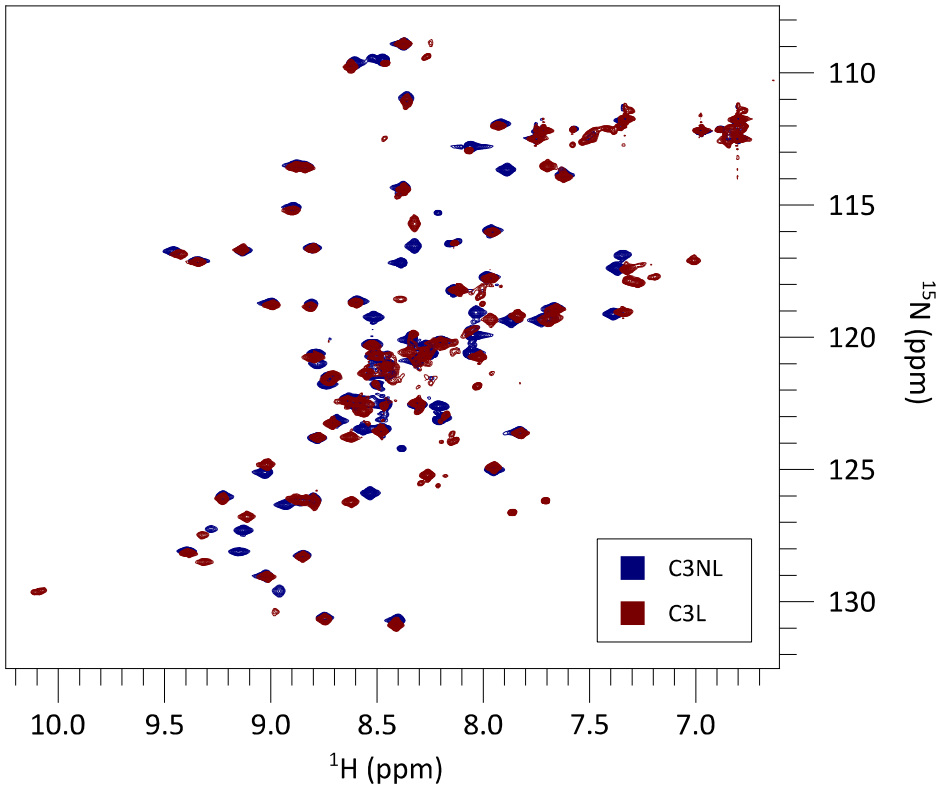


Figure 3.34: Overlay of the C3NL and C3L ^{15}N-HSQCs Displayed above is an overlay of the ^{15}N-HSQC spectra of the C3NL protein (blue), used for the structure in Figure 3.21, and C3L protein (red), expressed in Figure 3.33.

$$\Delta\delta = \sqrt{[(\Delta\delta_{1\mathrm{H}})^2 + (0.15\Delta\delta_{15\mathrm{N}})^2]}$$

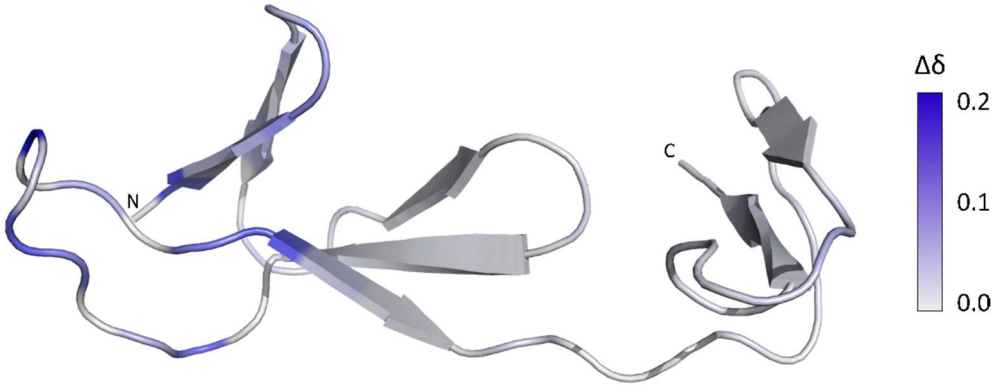


Figure 3.35: Backbone chemical shift changes between the C3NL and C3L proteins The structure below indicates the backbone $^{1}H_N{}^{15}N_H$ chemical shift changes (Δδ) throughout the protein, calculated by the equation displayed above[66]. Regions highlighted in blue indicate the greatest change in chemical shift while grey indicates no change in chemical shift.

The changes in chemical shift between the C3NL and C3L proteins (Figure 3.35) is centred around the N-terminal β-sheet and the long β4-to-β5 loop region, suggesting the N-terminal "tail" in the C3NL protein weakly interacts with this loop region. This indicates a potential change in conformation of this loop region in the WT C3 domain, however, the impact of appears to be minimal. The β4-to-β5 loop does not exhibit any obvious functional or structural properties relied upon by the domain as a whole, such as disulphide bonds, and the change in the N-terminal chain does not appear to affect the overall secondary and tertiary structure of the domain. Whilst future investigations would be useful to determine the precise orientation of the β4-to-β5 loop, this result does not affect the reliability of the structure presented in Figure 3.21 and the structural comparisons made to other VWFC domains.

4. Results: The effect of cysteine mutations on the structure of VWF C3

Although one of the least abundant amino acids in proteins making up only around 2.3% of amino acids in humans, cysteine is frequently observed in functionally important binding sites. Its presence in VWF is far higher than in human proteins on average, making up 8.3% of the protein[40]. It has unique properties, being able to react with another cysteine to form a disulphide bond whilst also being found as both hydrophobic and hydrophilic[426]. The importance of disulphide bonds in VWF is clear, allowing dimerisation and multimerisation to occur with the biological function of VWF relying on high molecular weight multimers[176]. Disulphide bonds are also relied upon in the functional binding sites of VWF, with vicinal disulphide bonds shown to be essential in the binding to platelet GpIbα and cleavage by ADAMTS13[294, 427]. The D and C domains are also rich in disulphide bonds, aiding stability with both domain types lacking a hydrophobic core[37]. The importance of free thiols on the physiological function of VWF is an emerging area of significance, with studies showing that blocking of free thiols impairs platelet recruitment under shear stress[267, 268, 277].

Although VWF free thiols, particularly in the C3 domain, appear to play a major role in lateral self-association having been found as free thiols in plasma-derived VWF[268, 274], the effect of these free thiols on the structure of the domain remains unclear. Therefore, we investigated the effects of mutating the two disulphide-bonded cysteines in each of the five C3 disulphide bonds to mimic disulphide bond reduction and used NMR to gain an initial insight into their effect on the C3 structure. Though mutations of two cysteines does not account for rearrangement of disulphide bonds by oxidoreductases in circulation, this begins to describe the effect on domain flexibility and maintenance of secondary and tertiary structure when disulphide bonds are reduced. Not only can this show the effects on the structural properties of the C3 domain, these mutations can display the importance of disulphides in forming a stable secondary and tertiary structure and indicate which disulphides in particular are heavily relied upon for the formation of the two individual subdomains.

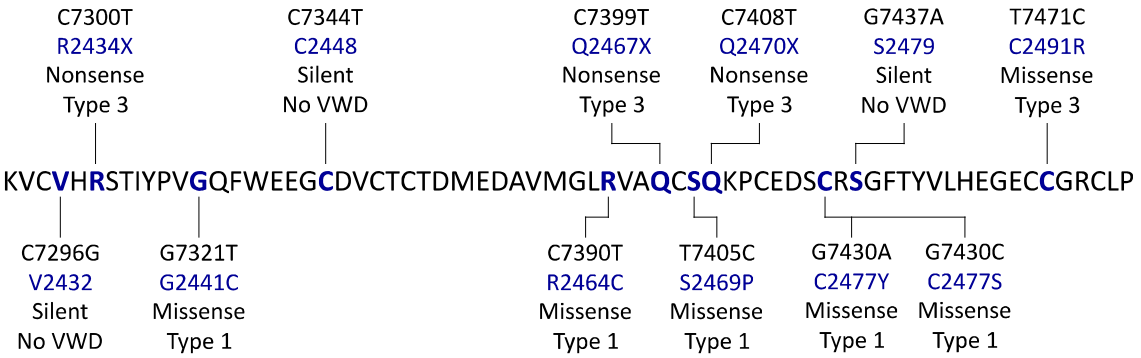


Figure 4.1: Locations of the VWD-causing substitution mutations in the C3 domain Displayed is the C3 domain sequence with those residues highlighted in blue being found in substitution mutations. For each residue where a mutation has been identified, the DNA mutation is displayed first, followed by the amino acid substitution. The type of mutation is displayed next, whether silent, missense or nonsense, and finally the VWD pathological result, whether Type 1, Type 3 or VWD is not seen.

Furthermore, mutations of the disulphide bonds can begin to explain the effect of VWD Type 1-causing mutations. There have been twelve substitution mutations in the C3 domain identified in patients (Figure 4.1), three of which, however, are silent mutations which as expected, do not result in VWD. Of those remaining, three are nonsense mutations which result in Type 3 VWD, importantly not expressing the crucial CK domain meaning that VWF fails to form dimers. There is also one mutation, C2491R, which interestingly results in Type 3 VWD. This may suggest a crucial role of C2491 in the proper folding of the C3 domain. The remaining five mutations each result in Type 1 VWD, four of which are caused by either a gain or loss of a cysteine residue. This suggests a crucial role of cysteines and disulphide bond formation in expressing properly folding VWF multimers. Whilst specific VWD Type 1-causing mutations were not expressed in this case, mutations of individual disulphide bonds can provide an insight into the impact of losing a disulphide bond on C3 domain folding. Expression of proteins with an odd number of cysteines may not be feasible in SHuffle *E. coli* cells[400] due to a high chance of protein misfolding in this case. Investigation of specific VWD Type 1-causing C3 mutations may, however, be performed in future studies, each of which are as a result of an odd number of C3 cysteines.

4.1. Mutations of the disulphide bonds of SD1

Double point mutations were performed on the C2431-C2453 and C2451-C2468 disulphide bonds. Note that these mutations were carried out on the same plasmid in which C3 was expressed for the structure calculation, not using the C3L plasmid. A substitution to either alanine or serine was considered, however, alanine was chosen due to its similar hydrophobic properties to cysteine. Mutations were carried out by PCR site-directed mutagenesis with primers designed to have minimal nucleotide changes. Primers were designed to be between 25 and 45 nucleotides in length in order for the primer to have sufficient binding affinity to the template plasmid. The T_m of the plasmid was to account for efficient annealing to the template plasmid, with the secondary structure $T_m < 42°C$ to ensure the secondary structure of the primer did not affect its annealing. For a full description of the PCR method, primers and thermocycling conditions, see Chapter 6.6. The PCR products were digested with Dpn1 and transformed into DH5α competent cells, ideal for high-efficiency cloning of PCR products.

Once each mutation was confirmed as successful, each of the C3 C2431A/C2453A and C2451A/C2468A mutants were expressed in M9 minimal media with $^{15}NH_4Cl$ and $^{12}C_6$-$^{1}H_{12}$-glucose, providing uniformly ^{15}N-labelled proteins to carry out NMR experiments on each mutant. The samples were purified in the same manner as in Chapter 2.7, using a Ni-NTA column to obtain a pure uncleaved sample and cleavage, reverse Ni-NTA and gel filtration performed as previously.

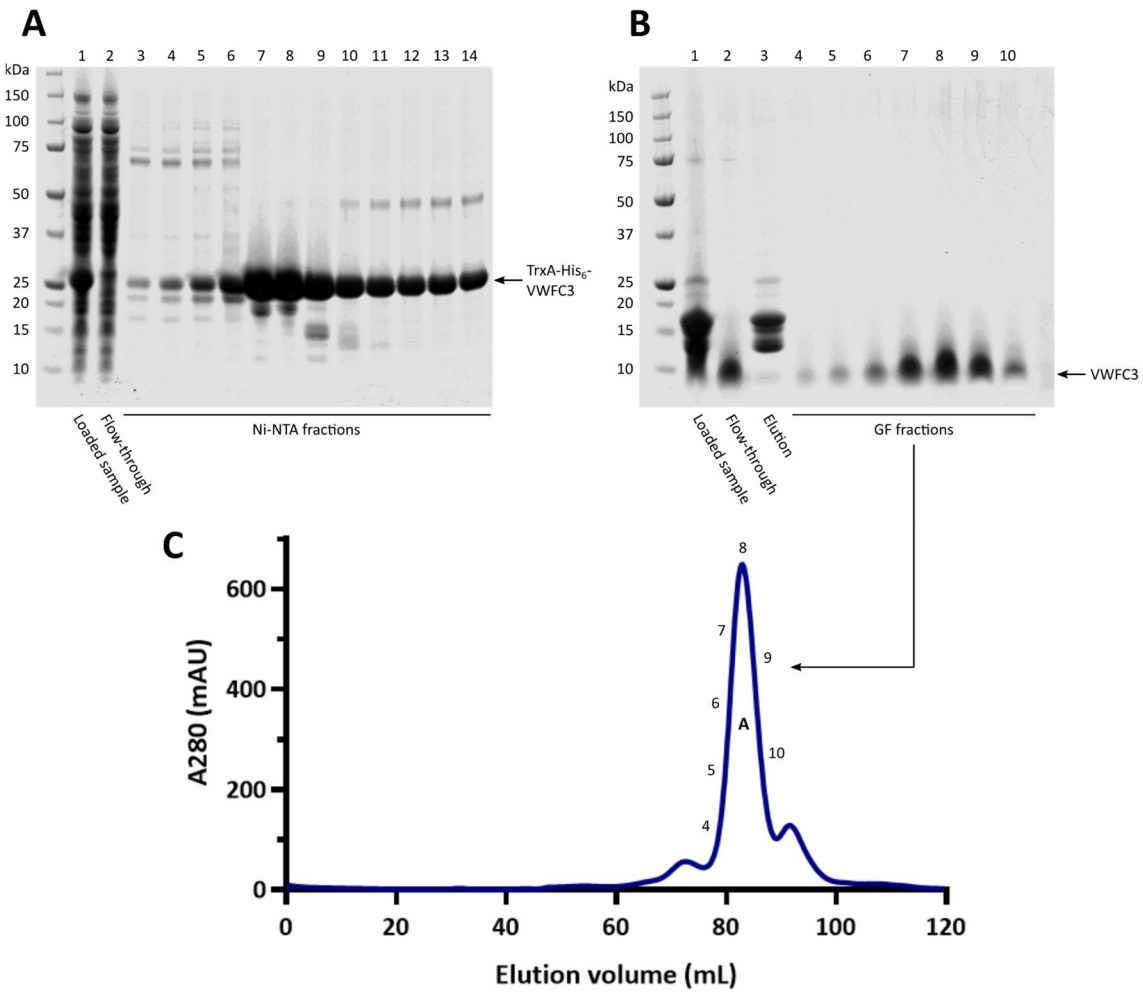


Figure 4.2: Purification of the VWF C3 C2431A/C2453A mutant The Ni-NTA affinity chromatography purification of the uncleaved TrxA-His_6-VWFC3(C2431A/C2453A) is shown in Panel A, analysing the elution fractions in reducing SDS-PAGE. Lane 1 shows the contents extracted from the cell cytoplasm, with Lane 2 showing the unbound protein that flowed through the column. Fractions eluted from the column are shown in Lanes 3-14. Panel B analyses the elution fractions from the size exclusion chromatography (Panel C), following a reverse Ni-NTA. Lane 1 shows the cleaved protein that was loaded onto the column, with the bound protein that was subsequently eluted from the column shown in Lane 3. The unbound protein in the flow-through (Lane 2) was separated by size exclusion chromatography. Lanes 4-10 correspond to the numbers seen on the chromatogram in Panel C, with the large band at ~8 kDa corresponding to cleaved VWF C3 (C2431A/C2453A).

The gel filtration profile and reducing SDS-PAGE results (Figure 4.2) show that the mutant protein was well expressed in M9 minimal media. Noticeably from both the SDS-PAGE of the gel filtration fractions and the elution profile is the difference in the apparent size of the C2431A/C2453A mutant compared to the WT C3 protein. The WT C3 protein presented at ~13 kDa in reducing SDS-PAGE and the elution peak centred at approximately 72.5 mL due to inaccurate cleavage. In the case of the C2431A/C2453A mutant, this displayed a size of ~8 kDa in reducing SDS-PAGE and the elution peak centred at 82.9 mL. This is similar to the protein expressed from the C3L plasmid which had accurate cleavage at the elution peak centred at

82.5 mL. The reason at this stage remained unclear as to why accurate cleavage appeared to occur in the case of the C2431A/C2453A mutant. The purification of the mutant C3 C2451A/C2468A protein is not displayed, however, displayed a similar purification result to that of the C2431A/C2453A mutant with the eluted gel filtration peak centred at 83 mL and the protein band presenting at ~8 kDa in reducing SDS-PAGE. The C2451A/C2468A mutant, however, did not express as efficiently as the C2431A/C2453A mutant. This is seen by a peak A280 reading of 640 mAU in the size exclusion chromatogram of the cleaved C2431A/C2453A mutant, compared to just 99 mAU in the case of C2451A/C2468A.

In order to determine any obvious effects of removing a disulphide bond on the structure of the VWF C3 domain, a ^{15}N-HSQC was carried out on each of the mutants.

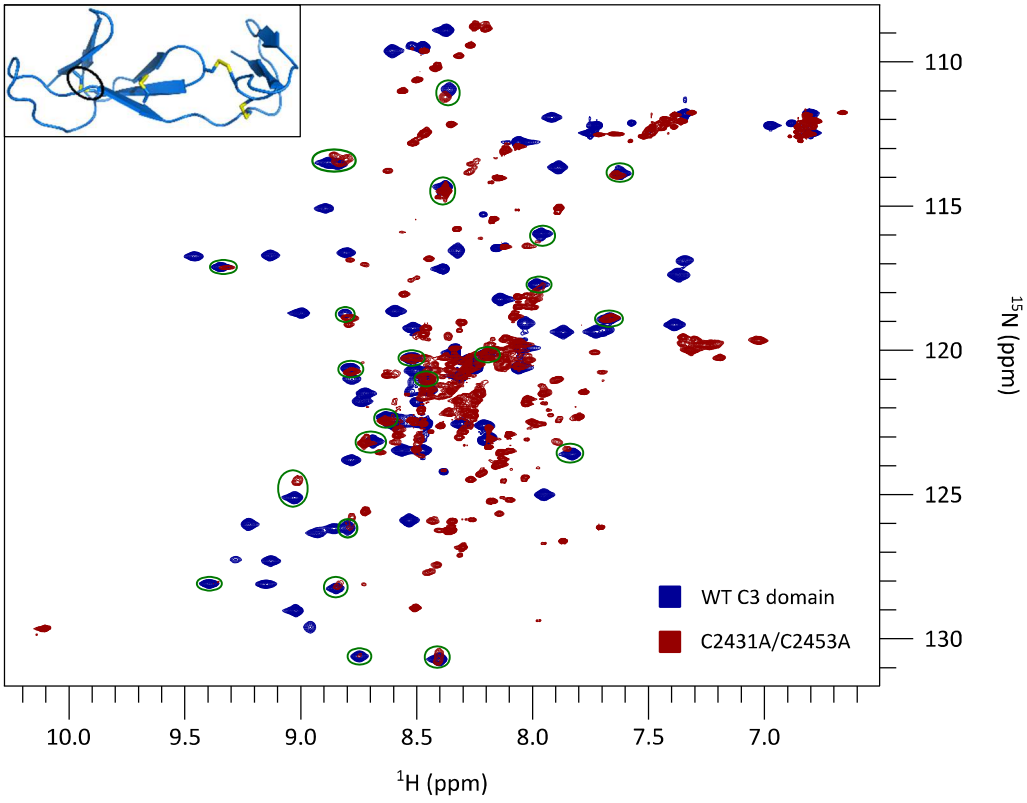


Figure 4.3: ^{15}N-HSQC of the C3 C2431A/C2453A mutant Displayed is an overlay of the ^{15}N-HSQC of the WT VWF C3 domain (blue) and overlaid is the ^{15}N-HSQC of the C3 C2431A/C2453A mutant (red). The mutated disulphide bond is circled in the structure shown in the upper left region of the spectrum. Circled in green are the peaks corresponding to SD2 of the WT C3 domain.

It is immediately apparent when observing the ^{15}N-HSQC of the C2431A/C2453A mutant (Figure 4.3) that the structure of the C3 domain is heavily compromised upon the removal of this disulphide bond. Peaks are concentrated in the ^{1}H 8-8.5 ppm range which indicates an unfolded protein but also rather broad, indicating that the protein is not completely disordered. This confirms a reliance by the C3 domain on disulphide bonds for a stable

structure, as was predicted from sequence analysis and the order parameter S^2 prediction by Talos+ (Figure 3.7). Particularly affected is the structure of SD1, with little to no overlap between peaks of WT C3 and the C2431A/C2453A mutant. The WT C3 peaks of SD2 do, however, overlay with peaks from the C2431A/C2453A mutant showing that SD2 still forms despite the removal of this disulphide bond. Although it is difficult to differentiate the SD2 peaks concentrated in the centre of the spectrum, the clear overlap of the well dispersed peaks allows this conclusion to be confidently drawn. This suggests that the two disulphide bonds in SD2, C2473-C2491 and C2477-C2494, are still able to form. Also noticeable is the increased number of peaks in the mutant spectrum, particularly in the upper portion corresponding to glycine residues. This suggests glycine residues, presumably from the unfolded SD1, reside in a range of conformations. The unfolding of SD1 also coincides with W2444 residing in its exposed conformation with the $^1H_\varepsilon{}^{15}N_\varepsilon$ peak of W2444 having a ^1H chemical shift of approximately 10.1 ppm. WT C3, on the other hand, favoured the closed conformation in around 90% of cases, as shown in Figure 3.14.

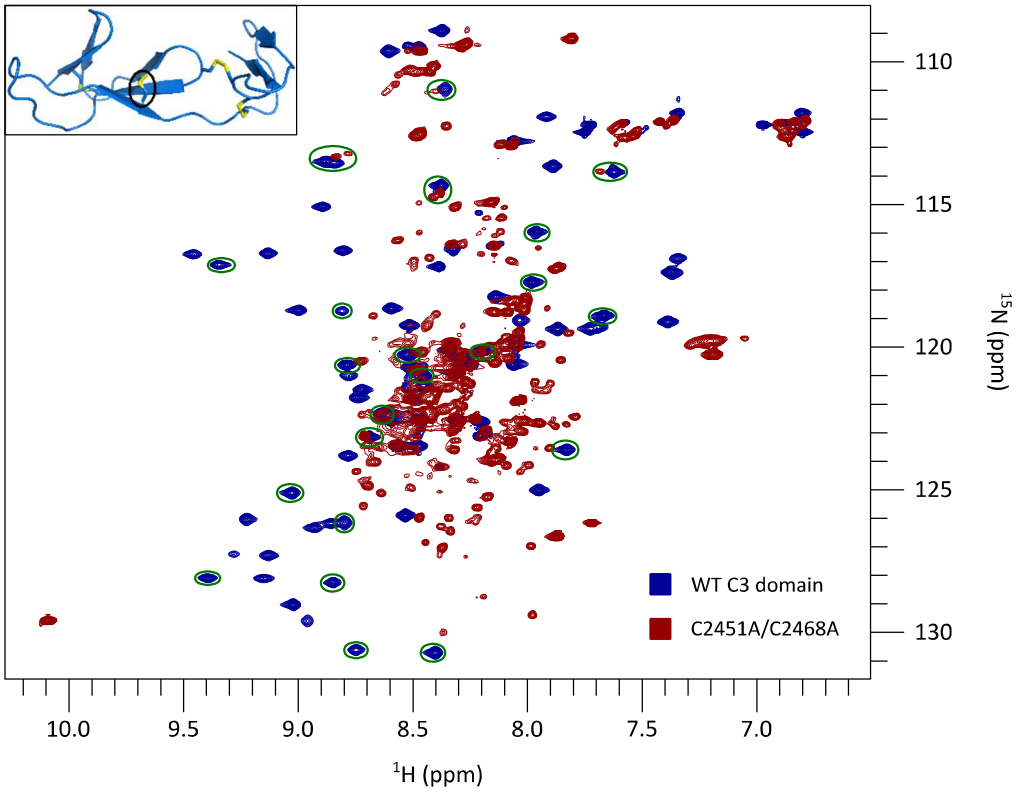


Figure 4.4: ^{15}N-HSQC of the C3 C2451A/C2468A mutant Displayed is an overlay of the ^{15}N-HSQC of the WT VWF C3 domain (blue) and overlaid is the ^{15}N-HSQC of the C3 C2451A/C2468A mutant (red). The mutated disulphide bond is circled in the structure shown in the upper left region of the spectrum. Circled in green are the peaks corresponding to SD2 of the WT C3 domain.

In the case of the C2451A/C2468A mutant (Figure 4.4), once again the secondary and tertiary structure of the domain appears to be lost with peaks concentrated in the 8-8.5 ppm range. There is little to no overlay with the peaks of SD1, indicating that its secondary and tertiary structure fails to form upon the removal of the C2451-C2468 disulphide bond. The overlay of the mutant ^{15}N-HSQC with WT C3 shows that SD2 does not fully form, however, certain residues do overlay potentially caused by partial folding of the subdomain. Whether partial folding is caused by the two disulphide bonds in SD2 continuing to form when the C2451A/C2468A mutant is introduced remains unclear. Compared to the C2431A/C2453A mutation, the introduction of the C2451A/C2468A mutation does not result in as high a number of glycine conformations despite its apparent SD1 unfolding, however, W2444 once again solely exists in its open conformation. Whilst it is difficult to analyse peak shift in the central region of the spectrum, it is clear that the mutation causes significant peak shift in 12 of the 19 well-dispersed SD2 peaks, showing that the subdomain does not form as in WT C3 (Figure 4.5). This highlights the impact of certain disulphide bonds even on neighbouring subdomains with SD2 evidently relying upon the C2451-C2468 disulphide for its proper folding and supports the idea of domain folding occurring in a cascade.

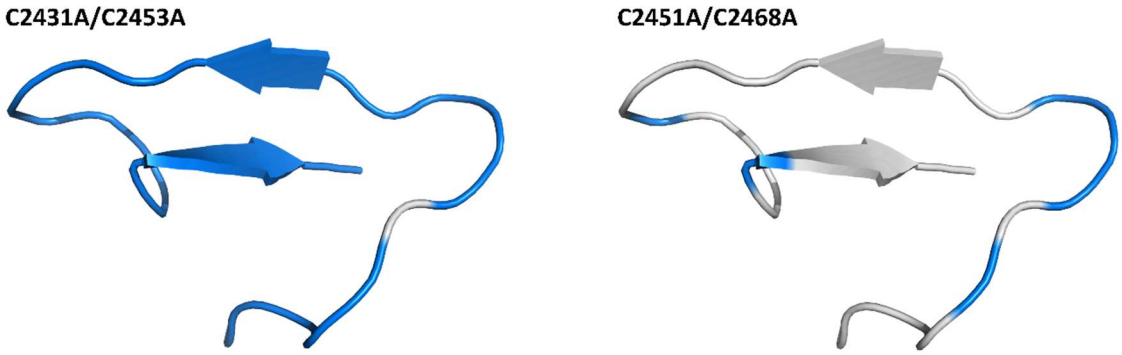


Figure 4.5: SD2 peak overlay in the SD1 disulphide bond mutants Each SD2 structure displayed indicates residues whose backbone amide ^{15}N-HSQC peak overlays with its corresponding peak in the WT C3 spectrum upon the introduction of a disulphide bond mutant (C2431A/C2453A or C2451A/C2468A). Regions highlighted in blue indicate peak overlay with minor peak shift whilst regions in white show a significant peak shift and lack of overlay. In both mutants, significant peak shift is observed throughout SD1.

Whilst analysis by a ^{15}N-HSQC does not provide a detailed description of how the structure changes or whether the other disulphide bonds continue to form despite the removal of their neighbouring disulphide, it does provide an insight into how the domain is affected by the reduction of a disulphide bond. The flexibility of the domain will be significantly affected by the reduction of either of the disulphides in SD1 having appeared to lose both secondary and tertiary structure. Relating these results to the effect of disulphide reduction on the C3 domain

in circulation, it will significantly increase flexibility of the domain and, hence, the VWF multimer. As discussed previously, a key role of the C domains is to add flexibility to VWF multimers due to domains being either heavily disulphide bonded, in the case of the D domains, or globular, the A domains. Therefore, adding flexibility will only aid the primary haemostatic function of VWF. Furthermore, greater flexibility will increase the likelihood of lateral self-association occurring under shear stress conditions, with a flexible domain allowing exposure of its free thiol.

4.2. Mutations of the disulphide bonds of SD2

Of the four cysteines that form the two intra-subdomain disulphide bonds of SD2, C2491 is the sole residue that is found as a free thiol in circulation. In order to determine the effect of reducing the C2473-C2491 disulphide bond on the structure of the C3 domain both C2491 and its disulphide-bonded cysteine, C2473, were mutated to alanine. The mutation to alanine was once again deemed most suitable to replace the cysteine residues due to their similar hydrophobicity. Although C2473 has not been identified as a free thiol in circulation, it was also deemed necessary to mutate this residue due to the likelihood of misfolding if the domain was expressed with an odd number of cysteines. This was seen previously where introduction of individual cysteine point mutations in C3 resulted in intracellular retention when expressed in full-length VWF[392].

Although neither cysteine in the second intra-subdomain disulphide bond in SD2, C2477-C2494, is found as a free thiol in circulation, a double point mutation of these cysteines to alanine would display any difference in unfolding between the reduction of C2473-C2491 and C2477-C2494. This may suggest a reason as to why C2491 is found as a free thiol, whereas, neither C2477 nor C2494 are. Primers for each of these mutations were designed using the same parameters as previously, with expression and purification of the mutated C3 protein being performed as was carried out for the WT C3 domain.

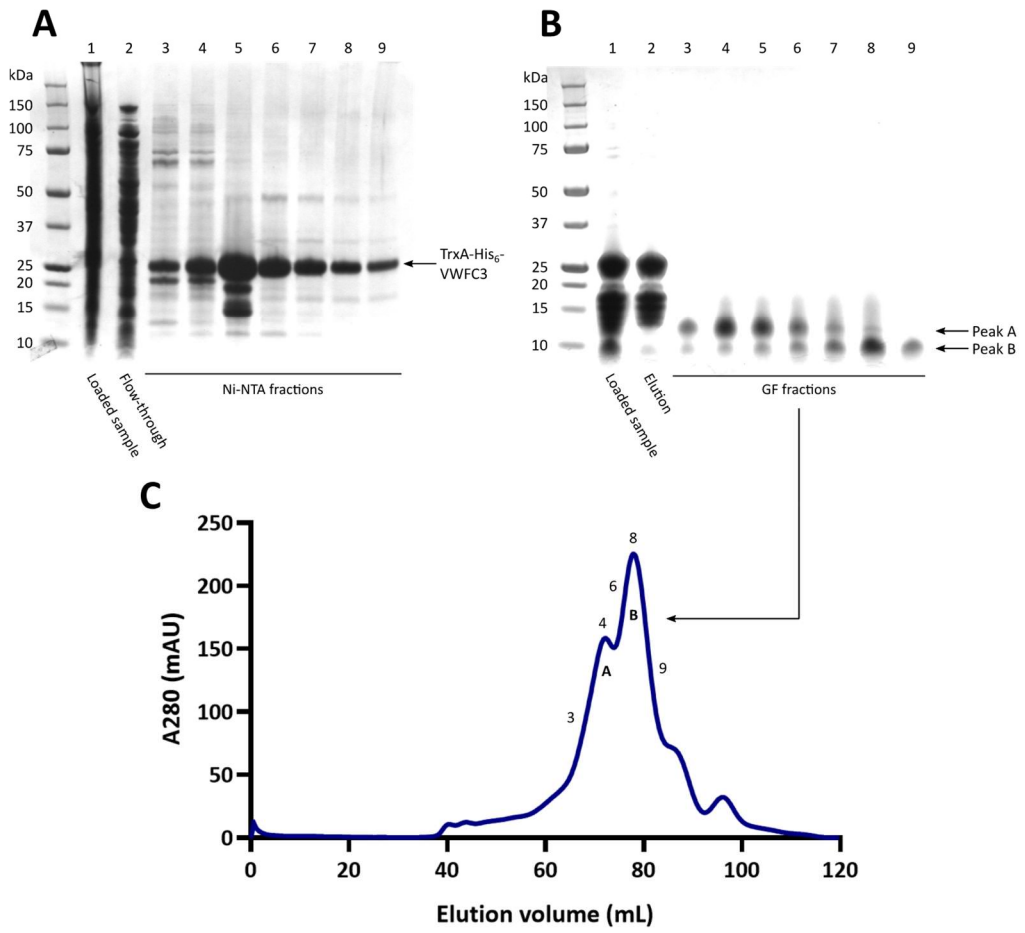


Figure 4.6: Purification of the C2477A/C2494A C3 mutant Panel A shows the Ni-NTA affinity chromatography purification of the uncleaved TrxA-His_6-VWFC3 (C2477A/C2494A), analysing the elution fractions in reducing SDS-PAGE. Lane 1 shows the contents extracted from the cell cytoplasm, with Lane 2 showing the unbound protein that flowed through the column. Fractions eluted from the column are shown in Lanes 3-9. Panel B analyses the elution fractions from size exclusion chromatography (Panel C), following a reverse Ni-NTA. Lane 1 shows the cleaved protein that was loaded onto the Ni-NTA column, with the bound protein that was subsequently eluted from the column shown in Lane 2. The unbound protein in the flow-through was separated by size exclusion chromatography and Lanes 3-9 correspond to the numbers seen on the chromatogram in Panel C. Also indicated on the chromatogram are the two distinct peaks, A and B.

The gel filtration trace of the cleaved C2477A/C2494A C3 mutant (Figure 4.6) displays two distinct peaks in the region where the C3 domain would be expected to elute. One peak was centred at approximately 72 mL and the second at 78 mL. Whilst not shown, the gel filtration trace of the C2473A/C2491A mutant also showed two peaks in identical positions. Similar to the purification of cleaved WT C3, the eluant at 72 mL suggests that cleavage had failed to occur at the intended site, however, the peak at 78 mL indicates some protein cleaving at the intended site. This is also shown in reducing SDS-PAGE with two distinct bands observed at ~13 kDa corresponding to the peak at 72 mL and ~9 kDa corresponding to the peak at 78 mL. Protein eluted in each peak was separately pooled, a ^{15}N-HSQC performed on both and the spectra compared (Figure 4.7).

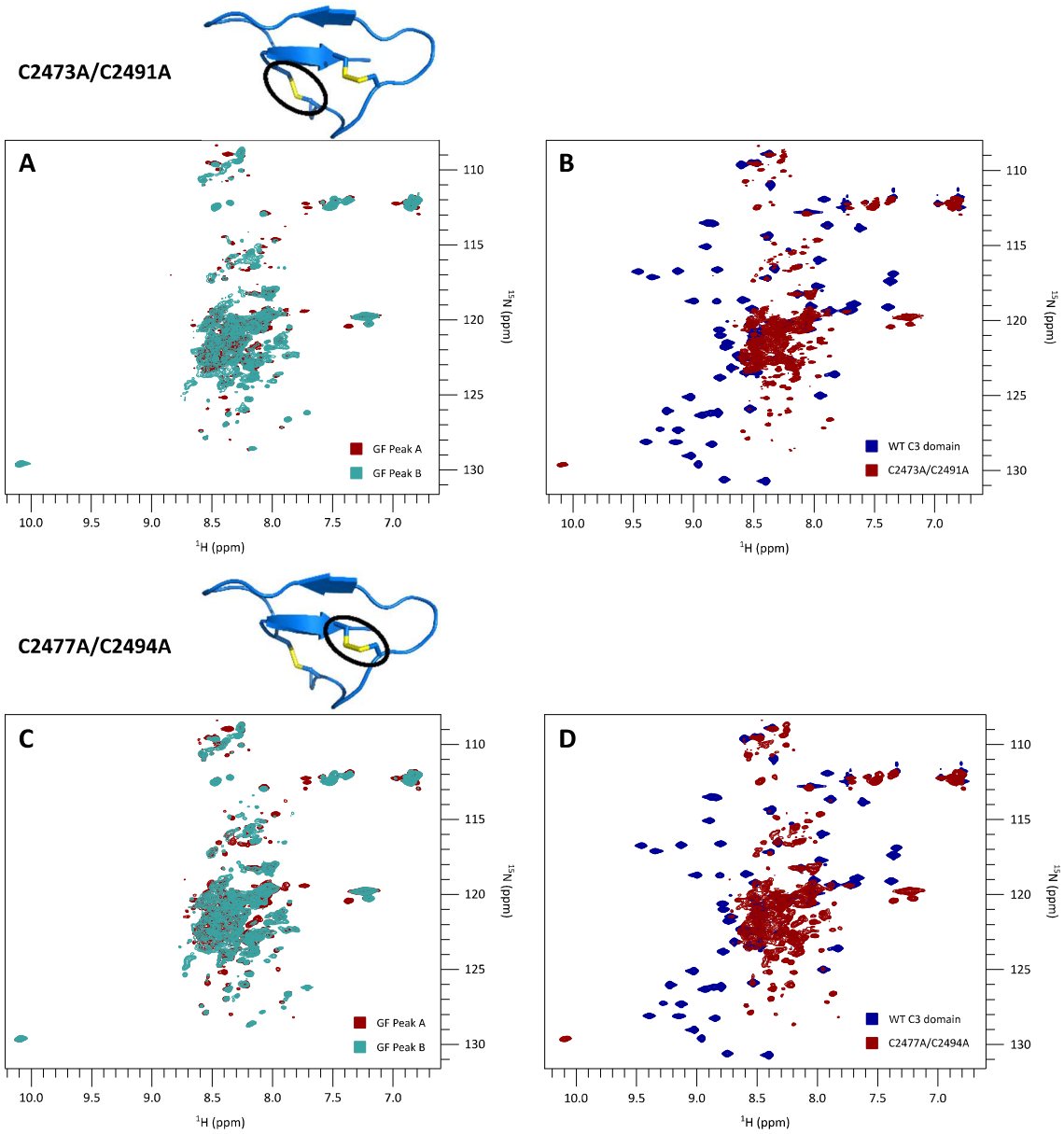


Figure 4.7: ^{15}N-HSQC of the C3 C2473A/C2491A and C2477A/C2494A mutants Displayed are the ^{15}N-HSQC overlays of the C2473A/C2491A mutant, panels A and B, and the C2477A/C2494A, panels C and D. Displayed in Panel A is the overlay of the protein eluted in Peaks A and B, denoted in Figure 4.6, with Peak A in red and Peak B in cyan. Panel B shows the overlay of the C3 C2473A/C2491A mutant protein, eluted in Peak A, with the WT C3 domain, shown in blue. Panels C and D show the equivalent spectra for the C2477A/C2494A mutant. The mutated disulphide bond in each case is highlighted in the displayed SD2 structure.

Comparison of the individual elution peaks, A and B, was performed to determine the similarity between the proteins eluted in each peak. It appears most likely that the two peaks occurs due to cleavage occurring in two separate positions, however, confirming the reason for this is not essential. In the case of the C2473A/C2491A C3 mutant, the comparison of the ^{15}N-HSQCs of the two eluted proteins shows similar spectra with peaks overlaying well and both spectra indicating poorly folded proteins. The same is true for the C2477A/C2494A C3

mutant with the proteins eluted in both peaks A and B displaying similar spectra and a lack of folding. When comparing to the WT C3 ^{15}N-HSQC, the protein eluted in peak A at 72 mL was used as a comparison due to the WT C3 protein also eluting at 72 mL, therefore, likely cleavage had occurred in identical places in both.

The overlay of the mutant C2473A/C2491A ^{15}N-HSQC with that of the WT (Figure 4.7, Panel B) shows a distinct lack of folding in the mutant C3 structure. Closer analysis indicates that there is no consistent overlay of peaks of either C3 subdomain. This is also observed in the ^{15}N-HSQC of the C2477A/C2494A mutant where the protein is unfolded and neither SD1 nor SD2 appear to have formed. The only obvious peak overlay is that of the side chain $H_\varepsilon N_\varepsilon$ of glutamine residues and $H_\varepsilon N_\varepsilon$ of W2444 which, in the mutant C3 proteins, is in the open conformation with a $^1H_\varepsilon$ chemical shift of ~10.1 ppm. This was also seen in the C2431A/C2453A and C2451A/C2468A mutants and further indicates a lack of folded structure in SD1. Interestingly, despite both mutants in this case being localised to SD2, SD1 fails to be folded. This suggests a reliance of SD1 on a properly folded SD2 structure to be able to form with an absence of factors within its own subdomain which would explain a lack of folding. Despite the lack of folding in SD1, it is unclear why cleavage did not solely occur at the intended EK cleavage site, as was the case with the mutants of SD1 disulphides when SD1 was unfolded. The results in this case, however, do not explain the reason as to why C2491 is found as a free thiol in circulation whereas the C2477-C2494 disulphide bond is not found reduced. Rather than this being due to a difference in the effect of reducing each disulphide bond, it may be explained by the C2473-C2491 disulphide having an increased susceptibility to reduction in circulation.

4.3. Mutation of the SD1-SD2 interconnecting disulphide bond

Both cysteines of the single inter-subdomain disulphide bond in VWF C3, C2448-C2490, are found as free thiols when in circulation[268]. The effect of reducing this disulphide is as yet unclear, however, with a lack of disulphide reduction within either subdomain, there is a greater probability that both subdomains remain folded as in the WT with increased flexibility between the two subdomains due to a lack of anchoring by the inter-subdomain disulphide bond. Whilst not confirmed whether both cysteines are found as free thiols simultaneously or if disulphide rearrangement occurs in some cases, in this case both cysteines were mutated to alanine and the effect on the C3 structure analysed.

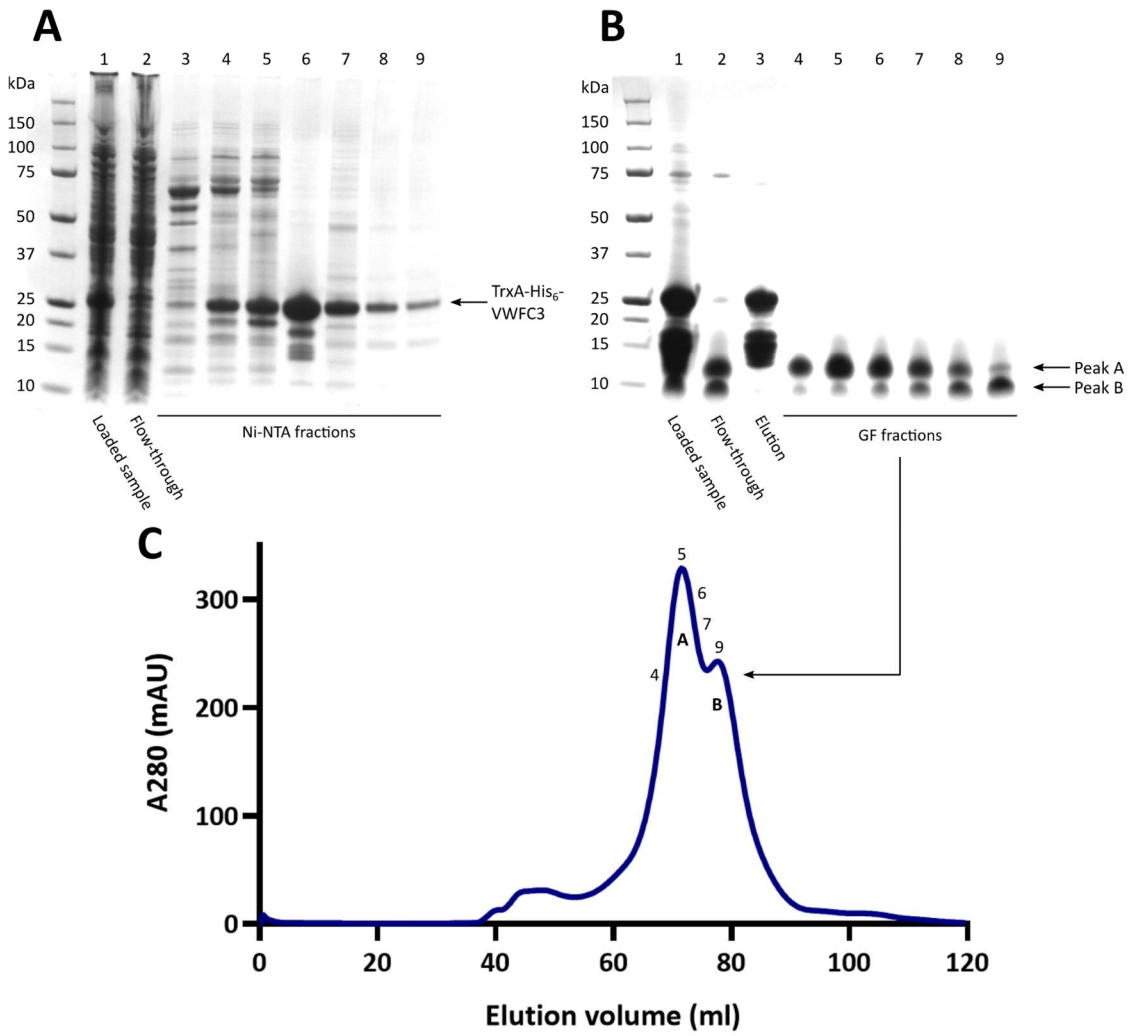


Figure 4.8: Purification of the C2448A/C2490A mutant The Ni-NTA affinity chromatography purification of the uncleaved TrxA-His_6-VWFC3 (C2448A/C2490A) is shown in Panel A, analysing the elution fractions in reducing SDS-PAGE. Lane 1 shows the contents extracted from the cell cytoplasm, with Lane 2 showing the unbound protein that flowed through the column. Fractions eluted from the column are shown in Lanes 3-9. Panel B analyses the elution fractions from size exclusion chromatography (Panel C), following a reverse Ni-NTA. Lane 1 shows the cleaved protein that was loaded onto the Ni-NTA column, with the bound protein that was subsequently eluted from the column shown in Lane 3. The unbound protein in the flow-through, shown in Lane 2, was separated by size exclusion chromatography and Lanes 4-9 correspond to the numbers seen on the chromatogram in Panel C. Also indicated on the chromatogram are the two distinct peaks, A and B.

As was seen in the purification of both the C2473A/C2491A and C2477A/C2494A mutants, the size exclusion chromatogram of the C2448A/C2490A mutant (Figure 4.8) presented two distinct peaks centred at approximately 72 mL and 78 mL, peaks A and B, respectively. These corresponded to protein bands in reducing SDS-PAGE of ~13 kDa and ~9 kDa, respectively. Similarly to the mutants of the SD2 disulphides, each peak was pooled separately, a ^{15}N-HSQC performed and the spectra compared (Figure 4.9). This would identify any difference in peak position between the two proteins or whether protein folding was identical between the two and the two peaks resulted from cleavage occurring at two individual sites.

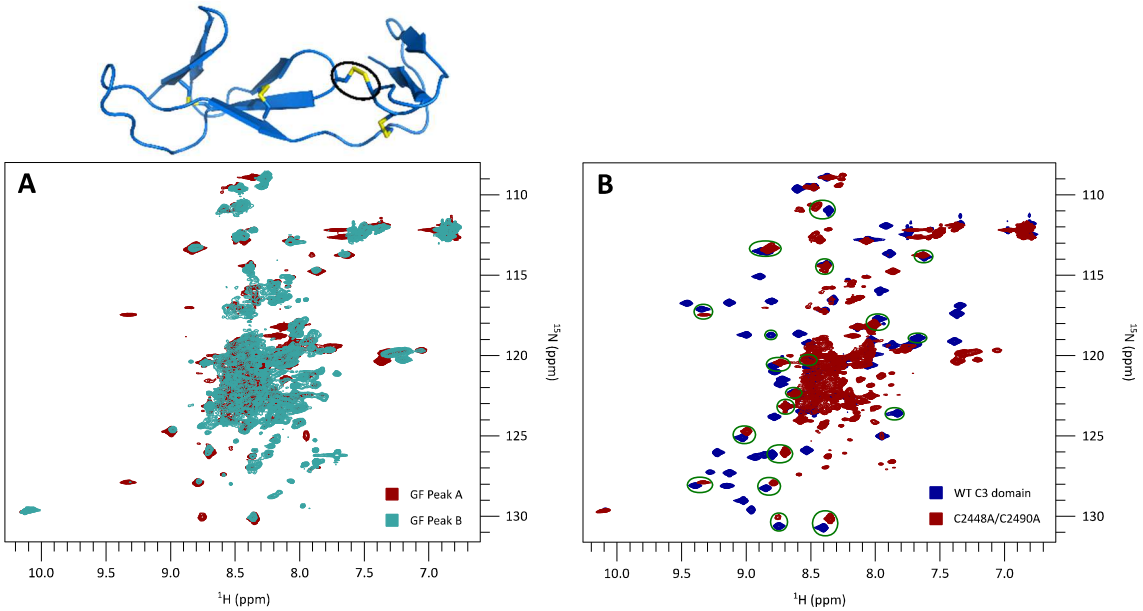


Figure 4.9: ^{15}N-HSQC of the C3 C2448A/C2490A mutant Displayed are the ^{15}N-HSQC overlays of the C2448A/C2490A mutant in panels A and B. Circled in the structure above is the mutated C2448-C2490 disulphide bond. Displayed in Panel A is the overlay of the protein eluted in Peaks A and B, denoted in Figure 4.8, with Peak A in red and Peak B in cyan. Panel B shows the overlay of the C3 C2448A/C2490A mutant protein, eluted in Peak A, with the WT C3 domain, shown in blue. Circled in green are the peaks corresponding to SD2 residues.

The overlay of the two distinct peaks, A and B, of the size exclusion chromatogram indicates that there is little difference in the spectra of peak A and B and, once again, cleavage had likely occurred at two separate sites. This includes the well dispersed peaks, however, the NMR signal of the protein from peak B was weaker. Therefore, when comparing the C2448A/C2490A mutant to WT C3, the protein eluted at 72 mL in peak A was used, also corresponding to the equivalent elution volume that was found in the purification of WT C3.

Comparison of the mutant C2448A/C2490A to WT C3 shows that peaks are concentrated in the central ^{1}H 8-8.5 ppm range, indicating a lack of folded structure. Peaks do, however, overlap well with the peaks of SD2 with minimal peak shift observed (summarised in Figure 4.10). This, once again, shows the reliance of SD1 folding on not just intra-subdomain interactions but also inter-subdomain interaction. The C2448A/C2490A mutant does not disturb any intra-subdomain interactions, however, there is an obvious reliance of SD1 on SD2 to either form or maintain a stable secondary or tertiary structure.

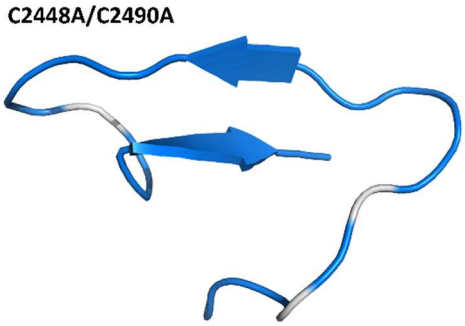


Figure 4.10: SD2 peak overlay in the SD1-SD2 disulphide bond mutant The WT SD2 structure displayed indicates residues whose backbone amide ^{15}N-HSQC peak overlays with its corresponding peak in the WT C3 spectrum upon the introduction of a disulphide bond mutant (C2448A/C2490A). Regions highlighted in blue indicate peak overlay with minor peak shift whilst regions in white show a significant peak shift and lack of overlay. Significant peak shift is observed throughout SD1.

In each of these mutants, how the subdomains unfold, whether the remaining disulphides form and whether certain parts of each subdomain continue to form remain uncertain, however, MD simulations may begin to provide an insight into the mechanism of unfolding. Initial MD simulations show that once the domain is formed and a disulphide bond is forcibly reduced, the structure of the C3 domain remains intact rather than the subdomains losing their secondary and tertiary structure, as is shown in the NMR spectra. Therefore, rather than the mutation of each disulphide bond rendering the tertiary structure of the C3 domain as wholly unstable, the domain is unable to fold initially as in the WT. Removal of one disulphide bond may cause a malformation of the remaining disulphide bonds, resulting in an unstable structure and presenting an NMR spectrum displaying an unfolded structure. Alternatively, removal of certain disulphide bonds may prevent the remaining disulphides from forming altogether. Additional MD simulations may provide further explanation for these results. The MD results indicate that reduction of certain disulphide bonds in circulation does not result in the total unfolding of the subdomains, rather an increase in flexibility at the localised region of the reduced disulphide bond. The mutations and NMR analysis performed, however, highlight the degree of protein misfolding as may occur in VWD Type 1-causing mutations. The sole mutations of a C3 cysteine that cause Type 1 VWD are C2477Y and C2477S, leaving a free C2494 residue. With the mutation of C2477A/C2494A resulting in a lack of folding across both subdomains, the chance of significant misfolding in the C2477Y and C2477S mutations is highly likely. Expression and NMR analysis of the C2477Y and C2477S mutants will provide further insight into their effects.

5. Concluding remarks

Since Dr Erik von Willebrand first identified a hereditary bleeding disorder in a series of consanguineous families in the Åland Islands in 1924, our understanding of VWF and its associated disorder has advanced immeasurably. The protein itself was first identified in 1971[17] and through laboratory studies and clinical observation, its unique characteristics have become clear. This extends to its multimeric synthesis, the multimeric structure itself and its haemostatic function.

The functional importance of VWF has become well characterised, acting as the first line of defence in the cessation of blood loss. It has a unique property of circulating freely in plasma in its globular form with its primary binding sites buried. Upon binding to exposed subendothelial collagen due to a vessel lesion and experiencing the local shear forces of laminar blood flow, the globular protein elongates, exposing its platelet binding sites[19]. Furthermore, the discovery of VWF's protective role of coagulation FVIII in circulation allowed the differentiation between haemophilia A and VWD Type 2N where inefficient VWF-FVIII binding causes reduced circulating levels of FVIII[21]. This is in contrast to haemophilia A where inefficient levels of circulating FVIII is due to compromised FVIII synthesis[428]. The array of VWF formations, whether stored in helical tubules, circulating as a globular protein or elongated in platelet capture, highlight versatile conformations of the multimeric protein. The dynamic nature of VWF function highlights the importance of domain structure and flexibility.

The understanding of VWF domain structure has improved drastically over recent years. This includes characterisation of the disulphide-bonded D domains, housing the major FVIII binding region of VWF, by NMR spectroscopy[62, 63] and the globular A domains by X-ray crystallography[88, 89, 293]. Although the structure of the C4 domain has been solved[125], investigation into the structure of the remaining VWF C domains has been largely unexplored and there is little knowledge regarding the structure of these domains. The structural significance of the C domains on multimeric VWF functionality is the flexibility within the domains and between the junctions, as is seen in other VWFC domains[49]. As a high resolution structure is limited to the C4 domain, subtle variations between the domains which aid multimeric flexibility remains unknown. These nuances between VWFC domains have been observed in other solved VWFC domain structures, such as a variation in the SD1-SD2 angle and the unique Clip segment in the VWFC domain of CV2. Whilst VWFC domains contain a number of common characteristics in disulphide bond arrangement and secondary structure,

they each display their own unique properties. In the case of the VWF C domains, it is essential to determine these variations and how they affect VWF functionality in health and disease.

The C3 structure calculated in this project shows significant similarities to the previously solved VWFC domain structures with two subdomains, seven strands of β-sheet and a disulphide bond arrangement identical to that which was predicted[37]. The C3 domain also exhibits its own unique properties with a long β4-to-β5 loop, something that is not seen in either the VWF C4 domain or the VWFC domain of CCN3 GF. The purpose of this unique β4-to-β5 loop as yet remains unclear, although with the 26-residue linker separating the C2 and C3 domains, it is possible that said loop interacts with the neighbouring C2 domain or rather occupies the space that results from the 26-residue separation between the two domains. The C3 domain also exhibits a unique orientation of SD2 with respect to SD1, being perpendicular due to the hydrogen bond between E2474 and R2478, rather than longitudinal as is observed in other VWFC domains with both the VWFC domain of CCN3 GF and C4 lacking the equivalent arginine residue. In terms of verifying the calculated C3 domain structure against the domain in WT full-length VWF, C3 does not contain a functional binding site. This means that a solid-phase binding assay to a known ligand, as was used in the C4 domain by binding to $\alpha_{IIb}\beta_3$[125], is not possible. The similarities, on the other hand, of the C3 domain to the VWFC domains and the C4 domain indicate the domain structure can be relied upon to represent that in WT VWF. Immediate plans, however, will be to update the calculated C3 structure using the protein expressed from the C3L plasmid, ensuring the conformation of the β1-β2 sheet and β4-to-β5 loop region are accurately described and not distorted by the linker impurity described in Chapter 2.8.

Questions may remain regarding the reliability of the C3 domain exhibiting the same folding as in WT VWF and whether the C3 domain folding occurs independently of the 2813 amino acid monomer. This method of protein expression, however, has been shown to be reliable in a number of other domain structures, including those of the VWF D', D3 and C4 domains[62, 63, 125]. When each of these domains were expressed individually, each resulted in efficient binding of ligands to their functional binding sites, suggesting that domain folding had occurred as in WT VWF. The fact that C3 was expressed in a similar manner and exhibited a similar folding pattern to C4 and other VWFC domains suggests that the method of protein expression can be relied upon. A comparison to other VWFC domains show similarities with regards to protein structure, particularly with regards to SD1 of the VWFC domains of both CCN3 GF and Col2A. Both of these domains interact with certain extracellular matrix proteins including BMP-2 and that of CCN3 GF interacts with transforming growth factor-β. No such interactions have yet

been identified in the VWF C domains, however, one common factor amongst all VWFC domains is significant flexibility in the linker that separates the two subdomains[47]. With each of these proteins displaying similar multimeric arrangements as that seen in VWF, their flexibility is likely essential to maintaining multimeric flexibility and physiological function.

Within VWF multimers, the C domains play a vital role in flexibility between monomers by their intra- and inter-domain flexibility. They significantly contribute to the flexibility and length that is vital for the transition between the globular and elongated states. The lack of such flexibility would affect the platelet binding capability of VWF under shear stress[126]. In terms of where the knowledge of the C3 domain structure can lead to, it acts as a starting point to understand a number of other aspects of the C domain physiology. The behaviour of the C3 domain under shear stress conditions can be investigated using laser tweezers[429], applying force to a single domain to test the flexibility between the two subdomains. Such a method has been used to investigate the unfolding and folding of the VWF A2 domain[288] and can provide the bridge from the NMR structure of the C3 domain to understanding its physiological significance in aiding multimeric flexibility. Whilst VWFC domain structures have been solved in CCN3 GF and Col2A[47, 49], the C3 domain is the first VWFC domain of VWF to have its structure presented with the C4 domain classed a VWFC-like domain[37]. This provides the basis to investigate the structures of the remaining VWF C domains. At this stage, it is unclear why the VWF C domains have diverged with only two high-resolution structures having been solved. One particular divergence between C3 and C4, however, is the β1-to-β2 loop which allows $\alpha_{IIb}\beta_3$ to efficiently bind to VWF. Once individual C domain structures are solved and the physiological significance of their characteristics are better understood, it may become clearer as to why divergence has occurred between these alike domains. The resolution of the individual C domains can also allow a progression onto the interactions of neighbouring domains in monomeric VWF. This would improve the understanding of how the C domains interact in circulating, globular VWF.

We undertook preliminary investigations into whether the C3 and C4 domain show any strong interaction, however, no obvious interaction was observed. In order to investigate this further, we would require a C3C4 construct and the interaction investigated by NMR. Expression of a C3C4 construct was performed, however, the issue of the construct comprising ten disulphide bonds meant that the concentration of the expressed protein was insufficient for analysis by NMR. The possibility that the expressed protein was retained in inclusion bodies, however, was not investigated. In the case that purification of protein retained within inclusion bodies also does not provide a sufficient protein concentration, a ligation of the two proteins may be

possible. Inaccurate cleavage on the C4 plasmid, however, may prove an issue in the case of ligating the two proteins. Even with a small linker as used in the C3L plasmid to improve cleavage specificity and accuracy, this would result in a short linker between the two domains whereas in WT VWF, C4 immediately follows C3 in the absence of a linker. More accurate cleavage may be overcome with the use of an alternative cleavage enzyme with TEV likely the most suitable candidate. The ligation of two neighbouring domains may be a useful technique, however, when investigating the interaction of the C2 and C3 domain, taking advantage of the 26 residue linker between C2 and C3. The design of a new C3 construct with said 26 residue linker separating the cleavage site and the C3 sequence would allow ligation to result in the identical sequence to that found in WT VWF.

The structure of the C3 domain can also provide the basis for understanding a number of other facets. Whilst this thesis displays the structure of the C3 domain in solution, additional microscopy studies can investigate the behaviour of the domain under shear stress conditions. Furthermore, the C3 domain structure can also lead to a greater understanding of the stem structure of the VWF dimeric bouquet when stored in WPBs. The C-terminal domains are closely associated in the stem of the VWF bouquet structure found in WPBs which may, in turn, interdigitate between adjacent VWF tubules, maintaining the spacing between tubules[195]. This structure has been visualised by negative stain EM[191], however, the underlying structural biology is as yet poorly understood. It is known that alike domains line up adjacent to each other in the stem structure in response to the pH drop from 7.4 to 6.2 between the ER and the Golgi and the increase in intracellular Ca^{2+} concentration[191]. It has also been shown that histidines in the D4 domain, the domain with the highest histidine density in the D4-CK region, promote strong intermonomer interactions below the critical pH of 6.8, leading to the closely associated dimers in the stem structure[430]. A detailed description, however, of the inter-domain stabilising forces between adjacent domains and the effect the WPB conditions have on the individual domain structures remains unexplained. With the C3 structure having been solved and the Golgi and WPB conditions well described, it may be possible to induce a C3-C3 dimeric structure as observed in the stem structure of the dimeric bouquet. This would allow the inter-domain stabilising forces to be characterised, identifying any changes in domain structure, whether disulphide bonds or tertiary stabilising interactions, and provide an insight into multimer rearrangement upon VWF secretion.

Whilst VWF disulphide bonds have been shown to be essential to its multimeric structure, free thiols in circulating VWF were also shown to be required for its full function under physiological flow conditions[267, 268, 277]. Evidence suggests a significant contribution of free thiol

groups to mediation of platelet adhesion under shear stress[267]. This has been shown through blockade of free thiols resulting in a virtual abolishment of platelet capture at shear rates above 5000 s^{-1}. This is likely explained by free thiol blocking reducing the flexibility of VWF multimers and its subsequent inability to withstand the high shear forces following collagen binding[277]. Although the effects of free thiol blocking have been well characterised, the underlying structural effects of disulphide bond reduction remain unexplored. This applies in particular to the C3 domain where five of the nine free thiol sites in VWF are found. With regards to investigating the structural effects of individual free thiols, it is not possible in an experimental sense to reduce and reoxidise individual disulphide bonds following protein synthesis, as occurs physiologically in circulation by TSP-1[261, 264]. This project, however, investigated the structural effect of reducing individual disulphide bonds, providing an insight into changes in domain structure and flexibility.

The effect of mutating each disulphide bond on the VWF tertiary structure is immediately apparent to see with each resulting in the loss of SD1 tertiary structure and only two of the five mutants (C2431A/C2453A and C2448A/C2490A) seemingly forming SD2 as in WT VWF. Whilst this clearly shows the increased flexibility of the domain and, in turn, VWF multimers upon the reduction of these disulphide bonds, this does not directly mimic the sequence of events that occurs in VWF synthesis and the reduction and rearrangement of disulphide bonds in circulation. Firstly, WT VWF is synthesised with each disulphide bond forming, however, the mutation of individual disulphide bond affects domain folding during protein synthesis. This difference is observed with the preliminary data obtained from MD simulations which showed the reduction of either disulphide bond in SD1 from the WT C3 domain did not result in the complete tertiary structure breakdown of SD1. The intricate process of protein synthesis and domain folding may be affected due to certain disulphide bonds relying on the formation of others to form. The mutation of individual disulphide bonds may affect the chain of disulphide bond formation. Furthermore, studies suggest not only reduction of disulphide bonds to form free thiols but also rearrangement of disulphide bonds[274]. Whilst this principle is not accounted for in the case of mutating individual disulphide bonds and protein synthesis may be affected, it does display the significant impact on domain structure and subsequent flexibility, particularly under shear stress. The use of MD simulations, however, will likely prove essential in investigating these structural effects, being able to reduce individual disulphide bonds following initial domain folding and subsequently induce rearrangement of certain disulphide bonds.

The C3 domain contains more VWD Type 1-causing mutations than any other C domain with each of these mutations either being of a cysteine or another residue to a cysteine[358]. Mutations causing Type 1 VWD largely cause issues regarding intracellular transport and lead to intracellular retention[355]. Whilst the disulphide bond mutations reported in this thesis do not directly correspond to VWD Type 1-causing mutations, they do highlight the extensive impact of removing a disulphide bond on domain folding. This would likely create issues in protein transport and storage, affecting the formation of the bouquet structure in which UL VWF multimers are stored and leading to intracellular retention of multimers. Whilst the MD simulations are a useful technique for investigating free thiol formation, they are less insightful in the case of VWD-causing mutations with the method focussing on a properly folded domain and the effect of reducing a single disulphide bond.

In summary, the NMR structure of the VWF C3 domain presented in this thesis is in agreement with the VWFC domains of other proteins in both its subdomain architecture and arrangement of its disulphide bonds. Important differences, however, between the C3 domain and other VWFC domains remain, such as the long β4-to-β5 loop and SD2 orientation, highlighting the importance of characterising the C3 domain structure. This provides an opportunity to gain a greater understanding of VWF storage and investigate the structural changes that occur between the pH-dependent stem structure and the circulating globular structure. With regards to the structural effects of VWF free thiols, this thesis indicates the extensive structural effects of reducing individual disulphide bonds and it provides a starting point in investigating this emerging and key facet of VWF physiology.

6. Materials and Methods

The C3 domain sequence was determined during the re-annotation of VWF where modules of VWF were assigned and related to the structure using EM[37]. A glycerol stock of a DNA plasmid in BL21 cells encoding for the VWF C3 domain was provided by Dr Thomas AJ Mckinnon (Imperial College London). The template plasmid was pET-32b(+) and, as discussed in Chapter 2.1, contained a TrxA tag, aiding the formation of the 5 disulphide bonds in C3, a His_6 tag, FXa cleavage site (Ile-Glu-Gly-Arg$^{\downarrow}$), the 68 residue C3 sequence and a second His_6 tag (TrxA-His_6-VWFC3-His_6). The expressed protein sequence is shown below.

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYG
IRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHHHHSSGLVPRGSGMKETAAAKFER
QHMDSPDLGTGGGSGIEGRKVCVHRSTIYPVGQFWEEGCDVCTCTDMEDAVMGLRVAQCSQKPCEDSC
RSGFTYVLHEGECCGRCLPGSNSPLAMAISDPNSSSVDKLAAALEHHHHHH

TrxA tag His$_6$ tag FXa cleavage site C3 protein

A Luria-Bertani (LB) Agar plate was made by dissolving 15g of Agar in LB Broth (25g LB powder, 1L water, autoclaved) and autoclaving. Into a petri dish was placed 20 mL of LB Agar (+100 µg/mL ampicillin) and left to set for 10 minutes. The TrxA-His_6-VWFC3-His_6 glycerol stock was streaked onto a LB Agar plate and incubated overnight at 37°C. A single colony was inoculated into 5 mL LB Broth and incubated at 37°C overnight. A subsequent mini prep was carried out to obtain a pure DNA sample which was sequenced by Source Bioscience and stored at -20°C.

6.1. LB expression and purification of TrxA-His_6-VWFC3-His_6

Expression in LB media was used to check the TrxA-His_6-VWFC3-His_6 protein could be successfully expressed, cleaved and purified before carrying out an expression using $^{13}C/^{15}N$-labelled isotopes. The TrxA-His_6-VWFC3-His_6 DNA vector was transformed into 50 µL NEB SHuffle® T7 competent *E. coli* cells[400]. A transformation was carried out by adding 2 µL DNA plasmid to 50 µL SHuffle *E. coli* cells and incubated on ice for 30 minutes. Cells were heat shocked at 42°C for 40 seconds and incubated on ice once again for 15 minutes. Added to the cells was 950 µL LB Broth and incubated at 37°C shaking for one hour. A LB Agar plate (+100 µg/mL ampicillin) was prepared and 50 µL transformed cells streaked onto the plate. Cells were incubated at 37°C overnight. A single colony was inoculated into 100 mL LB media (25 g LB powder/litre) and incubated at 37°C overnight in an orbital shaker (200 rpm) at 37°C. The following morning, a 2L LB Broth culture (+100 µg/mL ampicillin) was inoculated with the

overnight 100 mL culture. Throughout the project it was ensured that large cultures were inoculated with no more than 50 mL / litre culture as this would ensure that cells have sufficiently transitioned from the stationary phase reached overnight into the log phase. This ensures optimal cell growth and protein expression upon induction with IPTG. The 2L LB culture was grown at 37°C on an orbital shaker (200 rpm) to an OD_{600} of 0.65 with protein expression induced by adding 1 mM IPTG and incubating at 20°C for 18-20 hours. The cells were harvested by centrifugation at 4000 x g for 20 minutes at 4°C.

6.1.1. Protein Purification

Buffer	Chemical/preparation
Wash buffer	20 mM NaH_2PO_4 / Na_2HPO_4 500 mM NaCl 40 mM Imidazole pH 7.4, 0.22 µm filter sterilised, de-gassed
Elution buffer	Wash buffer + 500 mM Imidazole pH 7.4, 0.22 µm filter sterilised, de-gassed
Dialysis/gel filtration buffer	20 mM NaH_2PO_4 / Na_2HPO_4 100 mM NaCl pH 7.4

Table 6.1: Buffers used in the Ni-NTA and gel filtration purification

After buffers were prepared (Table 6.1), the cell pellet was resuspended in 30 mL of lysis buffer (30 mL wash buffer plus 3 mM $MgCl_2$, two tablets of Complete protease inhibitor cocktail (Roche), 10 µg/mL DNase, 10 µg/mL lysozyme) and lysed by high-speed vortex for two minutes and sonication (6x[30 seconds sonication, 30 seconds on ice]) until a smooth lysate had formed. The lysate was centrifuged at 39,000 x g for 60 minutes to remove the cellular debris and nuclei. The cleared cytoplasm was filtered through a 0.22 µm syringe filter. The protein was loaded onto a 5 mL His_6-Trap Ni-NTA column at 1 mL/min which had been pre-equilibrated with wash buffer. The column was washed with 5 column volumes of wash buffer and the column flow-through collected at each stage. The bound TrxA-His_6-VWFC3-His_6 protein was eluted using an AKTA fast protein liquid chromatography (FPLC) system (GE Healthcare) by a linear gradient of 40-400 mM Imidazole over 20 column volumes.

Eluted fractions were analysed in reducing conditions by SDS-PAGE alongside samples of the cleared cytoplasm that was loaded onto the column, the flow-through during the cytoplasm loading and flow-through at the wash step. Samples were set up using 15 µL protein sample (fractions with a high protein concentration were diluted 1 in 3 in wash buffer) + 5 µL loading

dye (4x) (200 mM Tris-HCl, 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol). Samples were denatured at 95°C for 4 minutes and loaded onto a NuPage 4-12% Bis-Tris Gel alongside 10 µL Precision Plus Protein standard (all SDS-PAGE gels described will be performed using this method, unless stated). Electrophoresis was performed for 45 minutes at 200 V. Gels were stained using InstantBlue™ Coomassie Protein Stain. Fractions containing the TrxA-His_6-VWFC3-His_6 protein were pooled and dialysed at 4°C overnight using a 10 kDa molecular weight cut off (MWCO) dialysis membrane against dialysis buffer.

6.1.2. FXa cleavage and purification

A small scale FXa cleavage of differing time points was set up (60 µL sample,2.5 µL FXa, 17.5 µL gel filtration buffer) at 4°C and 15 µL was taken at 2 hours, 4 hours, 6 hours, 18 hours and 24 hours and kept at -20°C. These samples were analysed using SDS-PAGE.

The remaining sample was cleaved (900 µL sample, 150 µL FXa, 1 mL gel filtration buffer) at 4°C for 24 hours. The cleaved sample was purified over a 1 mL Ni-NTA column pre-equilibrated with wash buffer and bound proteins eluted using elution buffer over a gradient of 40-500 mM imidazole. Eluted fractions were analysed by reducing SDS-PAGE with the result showing that the two fragments were unsuccessfully separated. Purification of the eluted flow-through was attempted over a gel filtration column (Superdex 75) pre-equilibrated with gel filtration buffer.

6.2. TrxA-His_6-VWFC3 (FXa) construct

6.2.1. Nonsense mutation of the TrxA-His_6-VWFC3-His_6 construct

As the VWFC3-His_6 cleaved protein was unsuccessfully separated from the TrxA-His_6 with either method, a PCR site-directed mutagenesis was performed to insert a termination codon before the second His_6 tag. This was carried out on the TrxA-His_6-VWFC3-His_6 plasmid, converting the glycine residue to a stop codon (GGC to TGA) immediately after the C3 sequence. The complementary mutagenesis oligonucleotides are displayed below with the mutation site in bold:

Forward:	5’	GGA AGG TGC CTG CCA **TGA** TCT AAC TCT CCT CTG G	3’	T_m: 65.6°C
Reverse:	5’	CCA GAG GAG AGT TAG A**TC A**TG GCA GGC ACC TTC C	3’	T_m: 63.6°C

The mutagenesis primer was designed to have a T_m below 68°C, the temperature of the annealing step. It was also ensured that the secondary structure of the primer had a T_m below 42°C, with any loop secondary structures within the primer affecting binding to the template

plasmid. Primers were synthesised by Sigma-Aldrich and delivered dried. Resuspension of primers was performed with Milli-Q® H_2O to be stored at a concentration of 100 µM.

The PCR was carried out using the Novagen KOD Xtreme™ Hot Start DNA Polymerase and the mutagenesis primers shown above. Primers were diluted 1 in 10 before addition to the PCR reaction, giving a stock concentration of 10 µM. The PCR reactions and thermocycling conditions were set up as Table 6.2 and Table 6.3, respectively.

	Stock	**Volume**	**Final concentration**
Xtreme Buffer	2x	25 µL	1x
dNTPs	2 mM	10 µL	0.4 mM
Forward primer	10 µM	2 µL	0.4 µM
Reverse primer	10 µM	2 µL	0.4 µM
DNA plasmid	238 ng/µL	1 µL	238 ng
Milli-Q® H_2O		9 µL	
DNA polymerase	1 U/µL	1 µL	1 U
Total volume		**50 µL**	

Table 6.2: PCR site-directed mutagenesis reaction setup

The thermocycling conditions were set up as below using a Bio-Rad T100™ Thermal Cycler. The extension step was calculated at 1 minute plus 1 minute per kb (DNA plasmid is 5.9 kb).

Step	**Temperature (°C)**	**Time (mins)**	**Cycles**
Polymerase activation	92	1:00	1
Denaturation	95	0:30	
Annealing	68	0:30	20
Extension	72	6:59	
Final extension	72	10:00	1
	4	∞	

Table 6.3: PCR site-directed mutagenesis thermocycling conditions

The PCR product was analysed using agarose gel electrophoresis against the template TrxA-His_6-VWFC3-His_6 DNA plasmid. Samples were set up with 5 µL DNA (PCR product/template plasmid) added to 1 µL of 6x Novagen DNA loading buffer. A 500 mL stock of 50x tris-acetate-EDTA (TAE) buffer was prepared using 121 g Tris Base, 28.5 mL glacial acetic acid, 100 mL 250 mM EDTA, pH 8.0. The agarose gel was prepared by adding 0.5 g agarose to 50 mL of 1x TAE buffer and heating. Once cooled, 3 µL of 10 mg/ml ethidium bromide was added, the

solution mixed and poured into a frame with a comb for well formation and allowed to cool at room temperature until set. Once set, the comb was removed and the agarose gel placed into a gel tank with 600 mL 1x TAE buffer added in order to cover the agarose gel with buffer. The template plasmid and PCR product were added to their individual wells and the gel run at 120 V for 30 minutes. A DNA ladder was not used as a comparison as the template plasmid was used to confirm a successful PCR reaction. Bands were visualised using ethidium bromide fluorescence under ultraviolet light.

Once the PCR reaction was confirmed successful, a DpnI digestion was carried out to digest the template plasmid (Table 6.4). Prokaryotes methylate their DNA at particular sequences to prevent them from being degraded by intracellular restriction enzymes. This is used as a defence mechanism against foreign DNA. Conversely, enzymes such as DpnI require DNA to be methylated at its restriction sites for it to be cleaved. In the case of a PCR reaction, the mutated plasmid produced in a PCR remains unmethylated due to the lack of methyltransferases as would be found in prokaryotes. This allows for the unmutated template plasmid, methylated from its production in *E. coli*, to be digested by DpnI, leaving solely the mutated PCR product[431].

	Volume (µL)
PCR product	20
Bovine serum albumin	0.5
NEB Buffer 4	2.5
DpnI	1
Milli-Q® H_2O	1
Total volume	**25**

Table 6.4: DpnI digest reaction setup

The 25 µL DpnI digest reaction was incubated for 2 hours at 37°C, following which the sample was incubated at 60°C for 30 minutes to deactivate the DpnI endonuclease and subsequently placed on ice. The DpnI reaction was transformed into Novagen NovaBlue™ cells, adding 8 µL of the DpnI reaction to 50 µL of cells. The cell transformation was performed using the same method as the TrxA-His_6-VWFC3-His_6 plasmid in Chapter 6.1. The only difference was following the addition of 950 µL LB Broth and incubation at 37°C, the 1 mL of transformed cells + LB Broth was centrifuged at 13,000 rpm for 5 minutes to pellet the cells. The supernatant was removed and cells resuspended in 200 µL of LB Broth. The 200 µL of resuspended cells were spread onto an agar plate (+ 100 µg/mL ampicillin) and incubated at 37°C for 16 hours until

colonies had formed. The transformation of a PCR product differs to that of an unmodified DNA plasmid due to the far lower concentration of a PCR product. By adding 8 µL of PCR product to the cells and streaking all cells onto the agar plate, this maximises the opportunity for successful colonies to form.

Into 5 mL of LB Broth cultures, four individual colonies were inoculated and grown in an orbital shaker at 37°C for 12-18 hours. The DNA of each culture was purified using a QIAgen Spin Miniprep kit and each purified DNA plasmid sequenced with a T7 forward and T7 reverse primer by Source Bioscience. This verified a successful mutation and a lack of PCR-mediated errors. This plasmid will be referred to as the C3 (FXa) plasmid. A successful mutant was transformed into SHuffle *E. coli* cells^{400} and a 2L LB expression, purification and dialysis of the uncleaved TrxA-His$_6$-VWFC3 was carried out as in Chapter 6.1. This provided a purified protein in the form as below.

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYG
IRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHHHHSSGLVPRGSGMKETAAAKFER
QHMDSPDLGTGGGSGIEGRKVCVHRSTIYPVGQFWEEGCDVCTCTDMEDAVMGLRVAQCSQKPCEDSC
RSGFTYVLHEGECCGRCLP

TrxA tag His$_6$ tag FXa cleavage site C3 protein

6.2.2. Expression of TrxA-His$_6$-VWFC3 in SHuffle *E. Coli* cells

The TrxA-His$_6$-VWFC3 plasmid was transformed into SHuffle® *E. coli* cells^{400} by the same method as that of the TrxA-His$_6$-VWFC3-His$_6$ plasmid (Chapter 6.1). The protein expression was also carried out by the same method as the of TrxA-His$_6$-VWFC3-His$_6$, however, using M9 minimal media instead of LB Broth (Table 6.5). The use of labelled ammonium chloride ($^{15}NH_4Cl$) and glucose ($^{13}C_6$-D-glucose) was dependent upon the experiments that were to be undertaken with the expressed protein sample. The added NH_4Cl and glucose was the sole source of nitrogen and carbon, respectively.

Buffer/Media	Chemical/preparation
M9 Base	53 mM $Na_2HPO_4.2H_2O$ 22 mM KH_2PO_4 8.6 mM NaCl pH 7.4
Micronutrient Solution 1 (MNS1)	14 g/L $CaCl_2.2H_2O$ 30 g/L Na_2EDTA 25 g/L $FeCl_3$
Micronutrient Solution 2 (MNS2)	24 g/L $CuSO_4.H_2O$ 18 g/L $MnSO_4.H_2O$ 25 g/L $CoCl_2$
M9 Media	M9 Base 0.1% MNS1 0.01% MNS2 1 mM $MgSO_4$ 0.1 mM $CaCl_2$ 0.01 g/L thiamine 0.01 g/L biotin *3 g/L glucose 1 g/L NH_4Cl

Table 6.5: M9 minimal media for expression of isotopically-labelled proteins *Glucose added at 10 g/L when using unlabelled $^{12}C_6$-$^{1}H_{12}$-glucose.

6.2.3. Purification of the VWF C3 protein

The cell pellet was harvested in the same method as that of TrxA-His_6-VWFC3-His_6, as was purification of the uncleaved protein using Ni-NTA affinity chromatography (Chapter 6.1). A 36-hour cleavage on a tube roller at 4°C was performed (2 ml sample, 1 mL gel filtration buffer, 300 µL FXa) and a reverse Ni-NTA pre-equilibrated with wash buffer to separate the cleaved C3 protein from any uncleaved protein and the TrxA-His_6 fragment. The column flow-through was purified by size-exclusion chromatography and the fractions analysed by SDS-PAGE. The TrxA-His_6 protein, separated from VWF C3 during the reverse Ni-NTA, was eluted off the Ni-NTA column with 100% elution buffer. The elution fraction was separated by size-exclusion chromatography to obtain a pure sample of TrxA-His_6.

The gel filtration peak centred at 72.5 mL of the C3 protein purification, expected to be the C3 protein, was pooled and concentrated in a 3 kDa MWCO to approximately 0.5 mL. Mass spectrometry analysis of the gel band believed to be VWF C3 was carried out by the BSRC Mass Spectrometry Facility at University of St Andrews. To the suspected C3 sample (in gel filtration buffer), pure TrxA-His_6 and an uncleaved TrxA-His_6-VWFC3 protein sample (both also concentrated to ~500 µL), 5% D2O was added and a ^{15}N-HSQC performed on each. The spectra

were overlaid to establish whether the sample obtained was the VWF C3 domain. Once this was confirmed by overlap of the ^{15}N-HSQC experiments and the mass spectrometry data, the VWF C3 sample was stored at room temperature for two weeks and a ^{15}N-HSQC carried out every three days to determine the stability of the protein at room temperature.

6.2.4. Ellman's assay

To establish free thiol concentration and whether all five C3 disulphide bonds had formed, an Ellman's assay was performed (Table 6.6). Ellman's reagent (DTNB) reacts with a free thiol group to produce a mixed disulphide bond and TNB, which can be detected spectrophotometrically at 412 nm.

Reaction buffer: 0.1 M sodium phosphate, pH 8.0, 1 mM EDTA

Ellman's reagent solution: Dissolved 4 mg Ellman's reagent in 1 mL of reaction buffer

Standard	Reaction buffer (mL)	Cysteine (MW = 175.6)	Final concentration (mM)
A	100	26.34 mg	1.5
B	5	25 mL of Standard A	1.25
C	10	20 mL of Standard A	1
D	15	15 mL of Standard A	0.75
E	20	10 mL of Standard A	0.5
G	25	5 mL of Standard A	0.25
F	30	0 mL	0

Table 6.6: Buffers and cysteine standards for an Ellman's assay preparation

A set of 8 tubes were prepared, one for each standard (A-F) and one for the C3 sample, each containing 50 µL Ellman's reagent solution and 2.5 mL reaction buffer. Added to each tube was 250 µL of the corresponding standard/unknown and were incubated at room temperature for 15 minutes. The absorbance was measured at 412 nm, a standard curve plotted, and the free thiol concentration of the unknown determined.

6.2.5. NMR experimental setup

NMR experiments were carried out with the help of Prof D. Flemming Hansen and Dr Angelo Figueiredo. Experiments were performed at 298 K on Bruker Avance III 700 MHz and 800 MHz spectrometers at the UCL NMR facility. Experiments were set up using parameters in Table 6.7.

Experiment	**Spectrometer (MHz)**	**Number of points (indirect dimension)**	**Spectral width (ppm)**	**Experiment time (hr:min:sec)**
^{15}N-HSQC	700	128 (^{15}N)	31.99 (^{15}N)	0:39:17
HNCO	700	42 (^{15}N) 162 (^{13}C)	35.00 (^{15}N) 14.00 (^{13}C)	37:04:57
HN(CA)CO	700	64 (^{15}N) 144 (^{13}C)	31.99 (^{15}N) 14.00 (^{13}C)	64:55:08
HNCACB	700	44 (^{15}N) 144 (^{13}C)	31.99 (^{15}N) 64.97 (^{13}C)	104:28:38
HN(CO)CACB	700	44 (^{15}N) 132 (^{13}C)	31.99 (^{15}N) 64.97 (^{13}C)	64:40:31
HNCA	700	44 (^{15}N) 132 (^{13}C)	31.99 (^{15}N) 30.01 (^{13}C)	63:07:54
HN(CO)CA	700	42 (^{15}N) 132 (^{13}C)	31.99 (^{15}N) 25.19 (^{13}C)	45:56:28
HCCH-TOCSY	800	132 (^{1}H) 6 (^{13}C)	5.87 (^{1}H) 69.99 (^{13}C)	66:37:51
CC(CO)NH-TOCSY	800	42 (^{15}N) 144 (^{13}C)	31.99 (^{15}N) 70.59 (^{13}C)	67:22:21

Table 6.7: NMR experiments for the backbone and side chain chemical shift assignment Displayed are the experiment details for each NMR experiment used for the backbone and side chain assignment. Details include the Bruker Avance III spectrometer that was used for each experiment, the number of points and spectral width in the indirect dimension and the total experiment time.

Spectra were processed in NMRPipe[432] and peak picking and further analysis performed using CCPN (**C**ollaborative **C**omputer **P**roject for **N**MR) software[433]. A backbone amide assignment was carried out determining the chemical shift of the backbone $^{1}H_{N}{}^{15}N_{H}$ on the ^{15}N-HSQC (one peak per amino acid, except for Proline due to a lack of a $^{1}H_{N}$). The triple resonance experiments – HNCO, HN(CA)CO, HNCACB, HN(CO)CACB, HNCA, HN(CO)CA[434, 435] – were used to assign the ^{15}N-HSQC peaks, supplemented with the CC(CO)NH-TOCSY experiment[436]. Backbone $^{1}H_{N}{}^{15}N_{H}$ chemical shifts were determined for 69% of residues due to a lack of signal which didn't allow for a completed sequential assignment. This resulted in a side chain chemical shift or NOE assignment not being performed. A linker impurity was also identified, originating from the linker between the FXa cleavage site and His_6 tag causing peak overlap in the triple resonance spectra.

6.2.6. Diffusion experiment

In order to determine whether the linker impurity was bound to the VWF C3 protein, a DOSY experiment was performed[409]. It was performed at 298 K on Bruker Avance III 800 MHz spectrometer with 14 points with a gradient strength of between 5% and 95% and the signal decay analysed using FuDA (**Fu**nction and **D**ata **A**nalysis). The diffusion coefficient of each assigned C3 and impurity peak calculated using the Stejskal-Tanner equation[437, 438] (Figure 6.1).

$$I = I_0 . e^{-\gamma_H^2 \delta^2 G^2 \sigma^2 . \left(\Delta - \frac{\delta}{3} - \frac{t}{2}\right) D}$$

I = Measured intensity

I_0 = Initial intensity

γ_H = Gyromagnetic ratio of hydrogen

δ = Duration of the pulse field

G = Gradient strength

σ = Structure factor

Δ = Time between the start of the two pulse field gradients

D = Diffusion coefficient

Figure 6.1: Stejskal-Tanner equation to determine the diffusion coefficient of a selected peak[437, 438]

After picking the peaks and calculating the diffusion coefficient, we determined that the C3 protein and impurity were bound to each other since they had the same diffusion coefficient.

6.3. TrxA-His_6-VWFC3 (EK) construct

6.3.1. New Plasmid design, protein expression and purification

Due to the difficulty in completing the NMR backbone assignment, a new plasmid was designed and provided by GenScript using the pET32b(+) vector as before. The redesigned plasmid, however, contained an EK cleavage site (Asp-Asp-Asp-Asp-Lys↓). This plasmid will be referred to as the C3 (EK) plasmid. The uncleaved protein sequence was as follows:

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYG
IRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHHHHSSGLVPRGSGMKETAAAKFER
QHMDSPDLGTDDDDKKVCVHRSTIYPVGQFWEEGCDVCTCTDMEDAVMGLRVAQCSQKPCEDSCRSGF
TYVLHEGECCGRCLP

TrxA tag His$_6$ tag EK cleavage site C3 protein

The transformation of the plasmid, protein expression in SHuffle *E. coli* cells[400] using M9 minimal media and uncleaved TrxA-His_6-VWFC3 purification carried out as with the C3 (FXa) plasmid (Chapter 6.2). After an initial ^{15}N-labelled protein expression to optimise cleavage conditions and verify the C3 protein using a ^{15}N-HSQC, isotopically-labelled VWF C3 was expressed in M9 minimal media and purified using the buffers described (Table 6.8).

Buffer	**Chemical/preparation**
Wash buffer	20 mM NaH_2PO_4 / Na_2HPO_4 500 mM NaCl 40 mM Imidazole pH 7.4, 0.22 µm filter sterilised, de-gassed
Elution buffer	Base buffer + 500 mM Imidazole pH 7.4, 0.22 µm filter sterilised, de-gassed
Dialysis buffer	20 mM Tris-HCl 100 mM NaCl pH 7.4
Wash buffer (post-cleavage)	20 mM Tris-HCl 100 mM NaCl 40 mM Imidazole pH 7.4, 0.22 µm filter sterilised
Elution buffer (post-cleavage)	20 mM Tris-HCl 100 mM NaCl 500 mM Imidazole pH 7.4, 0.22 µm filter sterilised
Gel filtration buffer	20 mM NaH_2PO_4 / Na_2HPO_4 50 mM NaCl pH 6.4, 0.22 µm filter sterilised, de-gassed

Table 6.8: Buffers used in the purification of the C3 protein

As opposed to the C3 purification (FXa plasmid), the dialysis buffer was Tris-based rather than phosphate due to inefficient cleavage of EK in phosphate buffer. Furthermore, the cleavage was carried out at 18°C in an orbital shaker (60 rpm) for 40 hours to optimise EK cleavage efficiency. The fact that the cleavage was performed in Tris buffer meant that the buffers used in the reverse Ni-NTA to purify the cleaved C3 protein were also Tris-based. Purification of the reverse Ni-NTA flow-through was also carried out as previously, over a gel filtration column (Superdex 75) pre-equilibrated with gel filtration buffer. Purified samples of VWF C3 were concentrated in a 3 kDa MWCO and NMR experiments performed with 5% D_2O, 2 mM NaN_3 and 2 mM EDTA.

6.3.2. NMR experiments and chemical shift assignment

NMR experiments were performed at 298 K on 600 MHz and 800 MHz spectrometers at the UCL NMR facility. Experiments were set up using the parameters described (Table 6.9). $^{13}C/^{15}N$-labelled C3 samples were at concentrations of 900 µM. A 700 µM fractionally-labelled C3 sample, using 10% $^{13}C_6$-$^{1}H_{12}$-D-glucose and 90% $^{12}C_6$-$^{1}H_{12}$-glucose, was also expressed for the stereospecific assignment of the methyl groups of valine and leucine.

Experiment	**Spectrometer (MHz)**	**Number of points (indirect dimensions)**	**Spectral width (ppm)**	**Experiment time (hr:min:sec)**
^{15}N-HSQC	600	256 (^{15}N)	32.03 (^{15}N)	1:20:16
HNCO	600	42 (^{15}N) 162 (^{13}C)	32.03 (^{15}N) 14.00 (^{13}C)	53:05:17
HN(CA)CO	600	42 (^{15}N) 162 (^{13}C)	32.03 (^{15}N) 14.00 (^{13}C)	53:29:40
HNCACB	600	44 (^{15}N) 144 (^{13}C)	32.03 (^{15}N) 64.99 (^{13}C)	70:40:09
HN(CO)CACB	600	44 (^{15}N) 132 (^{13}C)	32.03 (^{15}N) 64.99 (^{13}C)	65:36:49
CC(CO)NH-TOCSY	600	42 (^{15}N) 144 (^{13}C)	32.03 (^{15}N) 70.67 (^{13}C)	48:04:03
H(CC)(CO)NH-TOCSY	800	96 (^{15}N) 128 (^{1}H)	32.03 (^{15}N) 7.44 (^{1}H)	34:14:40
^{13}C-HSQC	800	256 (^{13}C)	37.99 (^{13}C)	0:36:45
HCCH-TOCSY	800	96 (^{1}H) 128 (^{13}C)	6.01 (^{1}H) 70.99 (^{13}C)	32:17:48

Table 6.9: NMR experiments for the backbone and side chain chemical shift assignment Displayed are the experiment details for each NMR experiment used for the backbone and side chain assignment. Details include the Bruker Avance III spectrometer that was used for each experiment, the number of points and spectral width in the indirect dimension and the total experiment time.

Spectra were processed in NMRPipe[432] with peak picking and the assignment carried out using CCPN software[433]. The backbone amide chemical shift assignment was annotated on the ^{15}N-HSQC, determined using the HNCO, HN(CA)CO, HNCACB and HN(CO)CACB, supplemented with the CC(CO)NH-TOCSY experiment[434-436] with a mixing time of 12 ms. The method of backbone amide assignment is described in detail in Chapter 2.5. The side chain chemical shift assignment, annotated on the ^{13}C-HSQC, was determined using the CC(CO)NH-TOCSY, H(CC)(CO)NH-TOCSY and HCCH-TOCSY experiments[439, 440]. The side chain chemical shift assignment method is described in Chapter 3.1. The assignment of prochiral methyl groups of

valine and leucine were also annotated on the ^{13}C-HSQC. A full list of the assigned C3 chemical shift is shown in Appendix 7.1.

6.3.3. Talos+ prediction of backbone dihedral angles and order parameter S^2

Backbone chemical shifts were input into Talos+[441], a protein backbone dihedral angle prediction program, to determine backbone torsion angles, ϕ and ψ. Dihedral angle predictions from loop regions of the C3 protein were excluded from the final structure calculation with constraints in loop regions reducing the reliability of the structure. The random coil index (RCI)-derived order parameter S^2 was also calculated by Talos+, providing information on backbone flexibility and, hence, the α-helix or β-sheet locations[442]. This prediction was used as a guide for the secondary structure assignment using NOESY spectra.

6.3.4. NOE assignment

In order to determine the secondary and tertiary structures, 3D NOESY experiments were carried out to identify neighbouring ^{1}H nuclei[443]. The parameters of the three NOESY experiments performed are displayed in Table 6.10.

Experiment	**Spectrometer (MHz)**	**Number of points (indirect dimension)**	**Spectral width (ppm)**	**Experiment time (hr:min:sec)**
^{15}N-NOESY-HSQC	600	96 (^{15}N) 256 (^{1}H)	24.87 (^{15}N) 11.20 (^{1}H)	71:48:49
^{13}C-NOESY-HSQC (aliphatic)	600	82 (^{13}C) 256 (^{1}H)	38.01 (^{13}C) 11.34 (^{1}H)	82:10:55
^{13}C-NOESY-HSQC (aromatic)	600	64 (^{13}C) 163 (^{1}H)	31.87 (^{13}C) 10.48 (^{1}H)	41:00:56

Table 6.10: NMR experiments for the NOE assignment Displayed are the experiment details for each NMR experiment used for the NOE assignment. Details include the Bruker Avance III spectrometer that was used for each experiment, the number of points and spectral width in the indirect dimension and the total experiment time.

Through-space interactions between neighbouring ^{1}H nuclei were identified using the three NOESY spectra, in turn determining the secondary and tertiary structure of the C3 domain. The NOE assignment provided a range of short, medium and long-range NOEs. Each ^{1}H-^{15}N spin-pair presented a diagonal peak in its corresponding strip of the ^{15}N-NOESY-HSQC with a series of NOEs in the form of cross-peaks. Each cross-peak corresponds to a neighbouring ^{1}H nucleus. Likewise, each strip in the ^{13}C-NOESY-HSQC displays a ^{1}H-^{13}C diagonal peak, with cross-peaks to neighbouring ^{1}H nuclei. A detailed description of the NOE assignment is shown in Chapters 3.2 and 3.3. Two separate ^{13}C-NOESY-HSQC spectra were performed, one each for

aliphatic and aromatic nuclei, due to the vastly different chemical shifts of aromatic and aliphatic ^{13}C nuclei. For example, aliphatic ^{13}C nuclei exhibited a chemical shift between 10 and 75 ppm, whilst aromatic ^{13}C nuclei were between 115 and 145 ppm. All NOEs were manually assigned, initially focussing on secondary structure NOEs, followed by longer range tertiary structure NOEs.

6.3.5. Residual dipolar couplings and relaxation experiments

Measurement of residual dipolar couplings were performed in the presence of a bacteriophage *Pf*1. This produces an anisotropic matrix which produces weak alignment of the protein in a certain orientation[417]. Conversely, in isotropic conditions in the absence of alignment media, each possible orientation has an equal probability. The measurement of RDCs was performed on an 800 MHz spectrometer with vector orientations calculated from the experimental residual dipolar couplings. A 2D in-phase/anti-phase (IPAP) ^{15}N-HSQC experiment was used to measure the $^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$ RDCs. The $^{13}C_{\alpha}^{i}$-$^{13}C_{\beta}^{i}$ RDCs were measured with a pair of ^{13}C-HSQC experiments where the $^{13}C_{\alpha}$-$^{13}C_{\beta}$ coupling was allowed to evolve or refocussed. Other RDC measurements were excluded from the final calculation due to unreliability in the final measurements. Backbone RDCs for C3 loop regions were also excluded from the final calculation because these do not appropriately report on the structure.

Flexibility of the C3 domain was assessed by ^{15}N relaxation experiments, measuring the longitudinal relaxation rate, R_1, and the relaxation rate in the rotating frame, $R_{1\rho}$. These experiments were recorded at 298 K on an 800 MHz spectrometer with peak intensities determined using FuDA[444]. Both R_1 and $R_{1\rho}$ were used to calculate the transverse relaxation rate, R_2, using the method outlined by Trott et al[445].

6.3.6. CYANA 3.0 structure calculation

A number of lists were included in the CYANA 3.0 structure calculation[419], exported from CCPN and Talos+. Chemical shift data and NOEs from the ^{13}C-NOESY (both aliphatic and aromatic) and ^{15}N-NOESY experiments were exported in a Sparky format. A number of commands were included in the calculation file to implement disulphide bond restraints, backbone dihedral angles, distance restraints between nuclei and RDC measurements (Figure 6.2). The five C3 disulphide bonds (C2431-C2453, C2448-C2490, C2451-C2468, C2473-C2491, C2477-C2494) were implemented using the *ssbond* command. The *read aco* command accounted for the *aco* file which included the backbone dihedral angles predicted by Talos+. The $^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$ and $^{13}C_{\alpha}^{i}$-$^{13}C_{\beta}^{i}$ RDCs were included in a separate *rdc* file with a *read rdc* command in the calculation file. The stereospecific methyl group assignment of the ten leucine and valine (V2430, V2432,

V2440, V2450, V2460, L2463, V2465, V2484, L2485 and L2495) residues were included with the *atom stereo* command in the initialisation file.

```
# Structure calculation with RDC data

weight_rdc    = 0.02                # weight for RDC restraints
cut_rdc       = 0.2                 # cutoff for RDC violation output
seed          = 274705              # random number generator seed

read upl c3.upl
ssbond  3-25
ssbond 20-62
ssbond 23-40
ssbond 45-63
ssbond 49-66
read aco c3.aco
read rdc c3.rdc

calc_all structures=100 steps=10000
overview c3.ovw structures=20 pdb
```

Figure 6.2: Calculation file for the CYANA structure calculation The calculation file displayed incorporates the *upl* file, detailing distance restraints between nuclei, the *aco* file, indicating backbone dihedral angles and the *rdc* file, with $^{1}H_{N}^{i}$-$^{15}N_{H}^{i}$ and $^{13}C_{\alpha}^{i}$-$^{13}C_{\beta}^{i}$ RDCs. Also included is the *ssbond* command, implementing the five disulphide bonds.

Using a *write upl* command in the initial CYANA calculation, a *upl* file was created detailing distance restraints between nuclei, both inter- and intra-residue. These were calculated from the input files. Each structure calculation created an overview *ovw* file, presenting a series of distance and torsion angle violations. Distance restraints that were violated were first confirmed and checked in the NOESY spectra and thereafter eased in the *upl* file. The calculated *pdb* file was input into the PSVS suite[421] to analyse torsion angle violations and restraints in the *aco* file were modified accordingly. Calculations and restraint modifications were repeated until minimal violations remained.

6.4. VWF C4 domain expression, purification and NMR analysis

A plasmid encoding for the VWF C4 domain was supplied by GenScript in the same form as that of the C3 domain. Displayed below is the sequence of the expressed protein.

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYG
IRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHHHHSSGLVPRGSGMKETAAAKFER
QHMDSPDLGTDDDDKSACEVVTGSPRGDSQSSWKSVGSQWASPENPCLINECVRVKEEVFIQQRNVSCP
QLEVPVCPSGFQLSCKTSACCPSCRCE

TrxA tag His$_6$ tag EK cleavage site C4 protein

The plasmid was transformed and expressed in the same method of that of the C3 domain (Chapter 6.3.1). That is, transformed into SHuffle *E. coli* cells[400], expressed in 2L of M9 minimal media, however, with unlabelled $^{12}C_6$-$^{1}H_{12}$-glucose. Purification and cleavage of the C4 domain were also identical to the C3 domain purification. The ^{15}N-HSQC experiment was performed on a 600 MHz spectrometer at 298 K in the same buffer as the C3 domain. The ^{15}N-labelled C3 and C4 domain were combined at equal concentrations and incubated at room temperature for 30 minutes prior to NMR analysis. The ^{15}N-HSQC to investigate the interaction of the C3 and C4 domain was also performed on a 600 MHz spectrometer. Both ^{15}N-HSQC experiments were performed with identical parameters to the ^{15}N-HSQC in the VWF C3 domain assignment (Table 6.9).

6.5. Expression, purification and NMR analysis of the C3L protein

The C3L plasmid was also supplied by GenScript, identical to the C3 (EK) plasmid, however, containing a four amino acid linker between the cleavage site and C3 protein sequence. The full sequence of the TrxA-His_6-VWFC3 protein is displayed below.

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYG
IRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHHHHHHSSGLVPRGSGMKETAAAKFER
QHMDSPDLGTDDDDKGSMAKVCVHRSTIYPVGQFWEEGCDVCTCTDMEDAVMGLRVAQCSQKPCEDS
CRSGFTYVLHEGECCGRCLP

TrxA tag His$_6$ tag EK cleavage site C3 protein

A ^{15}N-labelled C3 sample was expressed from the C3L plasmid using $^{15}NH_4Cl$ and unlabelled $^{12}C_6$-$^{1}H_{12}$-glucose. Expression in M9 minimal media and purification of the uncleaved protein were carried out in the same manner as that from the C3 (EK) plasmid. Cleavage was once again performed at 18°C in an orbital shaker (60 rpm), however, the cleavage time was only 18 hours due to the increased cleavage efficiency in this case. Purification of the cleaved protein was carried out as previously. A ^{15}N-HSQC was performed at 298 K on an 800 MHz spectrometer.

6.6. Disulphide bond mutations of the VWF C3 domain

Mutations of individual disulphide bonds of the C3 domain were performed on the C3 (EK) plasmid. Each disulphide bond (C2431-C2453, C2448-C2490, C2451-C2468, C2473-C2491 and C2477-C2494) was individually mutated to a pair of alanine residues, chosen due to its similar hydrophobic properties to cysteine. Mutations were performed on one cysteine which

comprises a disulphide bond and, once a successful mutation was confirmed through sequencing of the plasmid, its respective cysteine residue was mutated (i.e. C2431A mutation performed and confirmed, followed by the C2453A mutation). These mutations were performed by PCR site-directed mutagenesis with the Novagen KOD Xtreme™ Hot Start DNA Polymerase. Displayed below are the complementary mutagenesis oligonucleotides used for the cysteine to alanine mutations, with the mutation site highlighted in bold. The T_m for each primer is also displayed with primers designed to have a T_m below 68°C and a secondary structure T_m below 42°C.

C2431A

Forward:	GAC GAC GAC AAG AAA GTT **GC**C GTT CAT CGT AGC	T_m: 64.8°C
Reverse:	ATA GAT GGT GCT ACG ATG AAC G**GC** AAC TTT CTT GTC	T_m: 63.7°C

C2448A

Forward:	CTG GGA GGA AGG C**GC** CGA CGT TTG CAC	T_m: 66.9°C
Reverse:	AGG TGC AAA CGT CG**G C**GC CTT CCT CCC	T_m: 67.3°C

C2451A

Forward:	GCT GCG ACG TT**G C**CA CCT GCA CCG AC	T_m: 66.3°C
Reverse:	TCC ATG TCG GTG CAG GTG **GC**A ACG TCG	T_m: 66.3°C

C2453A

Forward:	GAC GTT TGC ACC **GC**C ACC GAC ATG GAA	T_m: 65.4°C
Reverse:	CGT CTT CCA TGT CGG TG**G C**GG TGC AAA C	T_m: 65.9°C

C2468A

Forward:	CTC CGT GTT GCG CAG **GC**C AGC CAA AAG	T_m: 66.2°C
Reverse:	GCT TTT GGC TG**G C**CT GCG CAA CAC GG	T_m: 66.9°C

C2473A

Forward:	CAG TCA AAA GCC G**GC** CGA GGA TAG CTG C	T_m: 64.1°C
Reverse:	GCA GCT ATC CTC G**GC** CGG CTT TTG ACT	T_m: 62.2°C

C2477A

Forward:	CCG TGC GAG GAT AGT **GC**C CGT AGC GGT TTT AC	T_m: 67.1°C
Reverse:	TAA AAC CGC TAC GG**G C**AC TAT CCT CGC ACG G	T_m: 66.7°C

C2490A

Forward:	CAT GAG GGC GAG **GC**C TGC GGT CGT TG	T_m: 67.0°C
Reverse:	GCA ACG ACC GCA G**GC** CTC GCC CTC AT	T_m: 67.5°C

C2491A

Forward:	AGG GCG AGT GC**G CT**G GTC GTT GCC TG	T_m: 66.8°C
Reverse:	AGG CAA CGA CC**A GC**G CAC TCG CCC TC	T_m: 66.1°C

C2494A

Forward:	GTG CTG TGG TCG T**GC** CCT GCC GTA AGC TG	T_m: 66.9°C
Reverse:	AGC TTA CGG CAG G**GC** ACG ACC ACA GCA CTC	T_m: 66.7°C

The reaction was set up as indicated (Table 6.11) with primers diluted 1 in 10, providing a stock concentration of 10 µM. The thermocycling conditions, carried out in a Bio-Rad T100™ Thermal Cycler, are displayed below. Compared to the mutagenesis in Chapter 6.2.1, the extension time was increased to 10 minutes to ensure full extension had taken place.

Step	Temperature (°C)	Time (mins)	Cycles
Polymerase activation	95	10:00	1
Denaturation	95	0:30	
Annealing	63	0:30	20
Extension	72	10:00	
Final extension	72	10:00	1
	4	∞	

Table 6.11: Thermocycling conditions of the PCR site-directed mutagenesis of C3 cysteine residues

A DpnI digest of the PCR reaction was performed to remove the template plasmid from the synthesised PCR product. The product was transformed into NEB DH5α high efficiency competent *E. coli* cells, with 8 µL of DpnI-digested product transformed into 50 µL cells.

Transformed cells were grown overnight on a LB Agar plate (+100 µg/ml ampicillin). Once colonies were grown, plasmid was mini-prepped and mutations were confirmed as successful, the mutated plasmid was transformed into SHuffle *E. coli* cells[400]. A 1L culture of each mutated protein was expressed in M9 minimal media (Table 6.5) with $^{15}NH_4Cl$ and unlabelled $^{12}C_6$-$^{1}H_{12}$-glucose to produce a ^{15}N-labelled sample. The protein, was cleaved and purified in the same manner as that described in Chapter 6.3.1. Gel filtration fractions were analysed in reducing SDS-PAGE and in the case where two peaks were observed in the trace, each peak was separately pooled and a ^{15}N-HSQC performed on both. The NMR experiments were performed on a Bruker Avance III 500 MHz spectrometer at 298 K. Each experiment was carried out with 512 points recorded in the indirect dimension and a spectral width of 35 ppm, giving an overall experiment time of 1hr 24 min 8 sec.

7. Appendix

7.1. VWF C3 chemical shift assignment

Displayed is the chemical shift data for the C3 domain in ppm. Blank spaces indicate the chemical shift was either not seen or left unassigned.

Residue	$^{1}H_{N}$	$^{1}H_{\alpha}$	$^{1}H_{\beta}$	$^{1}H_{\gamma}$	$^{1}H_{\delta}$	$^{1}H_{\varepsilon}$	$^{1}H_{\zeta}$	$^{13}C_{O}$	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	$^{13}C_{\gamma}$	$^{13}C_{\delta}$	$^{13}C_{\varepsilon}$	$^{13}C_{\zeta}$	$^{15}N_{H}$	$^{15}N_{\varepsilon}$
K2429	8.211	4.357	1.712 1.864	1.312 1.489	1.674 1.650	2.975		175.781	55.989	31.852	25.119	29.374	42.271		122.582	
V2430	7.886	5.133	2.106	0.649 0.689				174.518	59.042	36.381	17.495 21.408				113.654	
C2431	9.457	5.129	2.922 3.021					172.962	56.577	42.165					116.776	
V2432	8.781	4.854	1.889	0.840 0.761				175.695	61.113	33.141	21.420 20.757				121.006	
H2433	9.129	4.866	2.747 3.086		7.365	6.061		174.079	56.221	33.272		119.390	144.743		127.238	
R2434	9.283	3.712	1.803 1.833	0.855 1.280	3.098			176.460	57.793	27.238	27.370	43.533			127.238	
S2435	8.609	3.882	3.999 4.047					172.805	59.869	62.434					109.594	
T2436	7.863	4.273	3.935	0.578				171.195	61.898	70.388	20.380				119.326	
I2437	7.951	4.515	1.524	0.710 0.974 1.440	0.739			175.242	60.408	38.465	17.858 27.934	13.013			125.003	
Y2438	8.927	4.736	2.450 2.120		6.233	6.131		172.706	55.144	42.122		132.362	116.476		126.331	

Residue	$^{1}H_{N}$	$^{1}H_{\alpha}$	$^{1}H_{\beta}$	$^{1}H_{\gamma}$	$^{1}H_{\delta}$	$^{1}H_{\varepsilon}$	$^{1}H_{\zeta}$	$^{13}C_{O}$	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	$^{13}C_{\gamma}$	$^{13}C_{\delta}$	$^{13}C_{\varepsilon}$	$^{13}C_{\zeta}$	$^{15}N_{H}$	$^{15}N_{\varepsilon}$
P2439		4.793	2.370 1.958	2.014 2.069	3.696 3.743			176.704	62.029	32.547	27.457	50.625				
V2440	7.725	3.480	1.842	0.885 0.853				177.131	64.533	32.127	22.422 20.798				119.370	
G2441	8.893	3.655 4.399						174.025	44.788						115.108	
Q2442	7.696	4.568	2.285 2.255	2.585 2.635		6.975 7.723		175.108	55.840	30.395	34.481	179.491			119.326	112.168
F2443	8.597	5.560	2.876 2.936		7.125	7.261		175.409	56.486	42.168		132.314	131.842		122.408	
W2444	8.723	4.978	3.223 3.203		6.663	9.141 6.450	8.140 6.112	172.571	57.015	31.474		126.143	122.088	138.533 121.919	121.483	128.066
E2445	8.595	4.982	2.087 1.954	2.197 2.341				176.274	55.031	32.497	37.071				118.644	
E2446	8.304	4.643	1.952 2.146	2.362				175.793	55.146	31.873	36.124				122.478	
G2447	9.136	3.738 4.024						175.603	47.460						116.698	
C2448	9.228	4.959	2.787 3.663					173.728	53.492	36.771					126.021	
D2449	8.504	5.254	2.251 2.840					176.041	53.940	43.399					120.697	
V2450	8.996	4.480		0.878 0.848				175.288	60.971	33.706	20.978 21.075				118.720	
C2451	8.855	5.125	0.195 1.104					171.206	54.468	48.415					126.159	
T2452	7.918	4.514	3.369	0.789				172.328	59.508	70.896	18.503				111.941	
C2453	8.155	4.501	2.625 2.713					174.723	54.712	40.741					123.562	
T2454	8.501	4.607	4.364	1.205				171.697	61.586	71.046	21.790				121.272	

Residue	$^{1}H_{N}$	$^{1}H_{\alpha}$	$^{1}H_{\beta}$	$^{1}H_{\gamma}$	$^{1}H_{\delta}$	$^{1}H_{\varepsilon}$	$^{1}H_{\zeta}$	$^{13}C_{O}$	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	$^{13}C_{\gamma}$	$^{13}C_{\delta}$	$^{13}C_{\varepsilon}$	$^{13}C_{\zeta}$	$^{15}N_{H}$	$^{15}N_{\varepsilon}$
D2455	8.321	4.674	2.575 2.690					176.722	53.712	41.303					116.500	
M2456	8.062	4.400	1.983 1.993	2.495 2.573				175.332	55.556	33.516	32.100				120.630	
E2457	8.594	4.281	1.877 1.915	2.135 2.180				176.214	56.604	31.026	36.409				122.565	
D2458	8.566	4.555	2.529 2.701					176.245	53.893	42.273					123.487	
A2459	8.533	4.191	1.433					178.312	53.919	19.153					125.905	
V2460	8.140	3.948	2.149	0.976 0.932				177.058	64.078	32.444	21.523 21.165				118.191	
M2461	8.515	4.432	2.549 2.676	2.050 2.104				177.212	55.934	32.397	32.969				119.254	
G2462	8.371	3.861 4.065						173.720	45.990						108.906	
L2463	7.389	4.814	1.687 1.850	0.920	0.910			176.350	54.068	44.193	25.448	23.709			119.132	
R2464	8.453	4.914	2.256 1.777	1.613 1.624	3.006 3.122	7.374		175.978	55.116	31.107	28.387	43.301			121.344	85.243
V2465	8.036	4.458	1.984	0.904 0.955				176.080	61.546	34.535	20.944 21.381				119.035	
A2466	8.956	5.120	1.305					176.299	51.295	20.439					129.639	
Q2467	8.740	4.601	1.784 1.902	2.239 2.164		7.341 6.806		174.861	54.739	31.512	33.761	180.662			111.918	111.784
C2468	8.781	5.391	2.917 2.589					173.389	55.326	48.342					123.831	
S2469	8.803	4.598	3.692 3.778					172.418	57.079	65.627					116.641	
Q2470	8.475	4.603	1.889 2.017	2.462 2.374		6.800 7.752		177.134	56.867	29.195	33.953				123.470	112.480

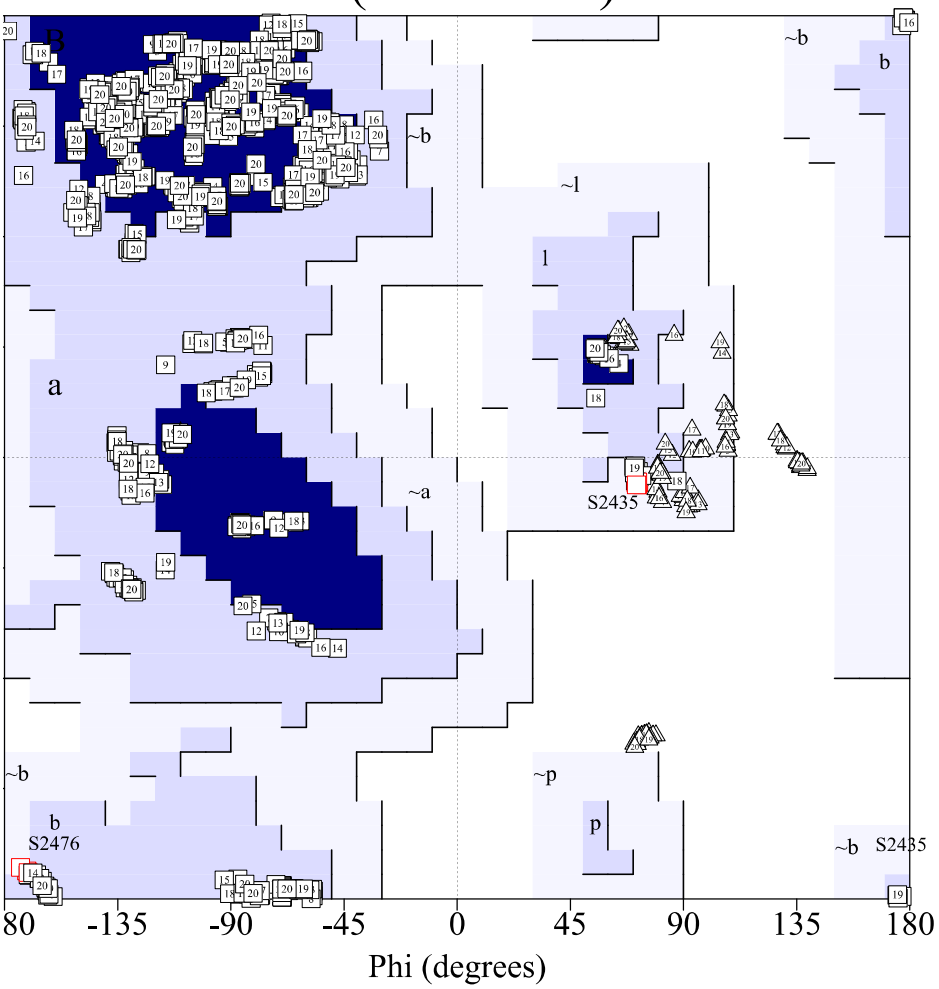
Residue	$^{1}H_{N}$	$^{1}H_{\alpha}$	$^{1}H_{\beta}$	$^{1}H_{\gamma}$	$^{1}H_{\delta}$	$^{1}H_{\varepsilon}$	$^{1}H_{\zeta}$	$^{13}C_{O}$	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	$^{13}C_{\gamma}$	$^{13}C_{\delta}$	$^{13}C_{\varepsilon}$	$^{13}C_{\zeta}$	$^{15}N_{H}$	$^{15}N_{\varepsilon}$
K2471	9.019	4.289	1.791 1.656	1.479 1.345	1.573	2.868		174.525	55.475	32.552	25.238	29.898	41.944		129.022	
P2472		4.377	1.841 2.247	1.957 2.030	3.593 3.869			176.244	63.107	32.237	27.453	50.859				
C2473	8.524	4.952	3.044 2.777					174.570	52.950	38.478					120.296	
E2474	8.798	4.380	1.932 2.068	2.219 2.237				175.017	56.011	28.835	36.282				126.159	
D2475	7.828	3.985	1.780 2.005					174.971	52.841	39.879					123.597	
S2476	7.621	4.445	3.742 3.760					174.198	56.925	63.556					113.881	
C2477	8.452	4.914	3.158 2.614					175.164	52.443	40.094					121.014	
R2478	8.636	4.287	1.845 1.622	1.556	3.007			175.561	55.823	30.980	27.075	43.394			122.327	
S2479	8.373	4.228	3.860 3.844					175.637	60.211	62.937					114.288	
G2480	8.841	3.576 4.232						173.937	44.987						113.534	
F2481	8.201	4.901	3.286 2.624	7.313	6.875			174.881	56.502	41.221		131.237	131.215		120.152	
T2482	9.348	4.508	3.942	1.167				172.815	61.799	71.477	21.326				117.099	
Y2483	8.847	4.689	2.904 3.038		6.915	6.656		174.659	58.557	38.485		133.047	118.355		128.258	
V2484	8.751	3.974	1.766	0.882 0.595				173.993	61.886	34.151	21.054 20.679				130.676	
L2485	8.400	4.292	1.610 1.066	1.331	0.863 0.784			175.602	53.994	44.636	27.054	24.338 25.347			130.724	

Residue	$^{1}H_{N}$	$^{1}H_{\alpha}$	$^{1}H_{\beta}$	$^{1}H_{\gamma}$	$^{1}H_{\delta}$	$^{1}H_{\varepsilon}$	$^{1}H_{\zeta}$	$^{13}C_{O}$	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	$^{13}C_{\gamma}$	$^{13}C_{\delta}$	$^{13}C_{\varepsilon}$	$^{13}C_{\zeta}$	$^{15}N_{H}$	$^{15}N_{\varepsilon}$
H2486	9.026	4.777	2.824 3.197		7.155	8.084		174.234	54.294	30.601		120.178	136.399		125.072	
E2487	8.694	4.050	1.947 1.986	2.271				177.441	58.288	29.433	36.011				123.168	
G2488	8.879	3.673 4.234						174.212	45.329						113.522	
E2489	7.666	4.563	2.185 2.044	2.222				176.820	55.271	31.004	36.803				118.869	
C2490	8.811	4.560	3.237 2.864					175.147	56.525	37.835					118.774	
C2491	7.963	4.985	2.833 3.225					173.946	52.536	40.179					115.972	
G2492	8.357	2.960 4.376						172.304	45.225						111.013	
R2493	7.982	4.764	1.605 1.792	1.607 1.603	3.179 3.123			174.536	54.309	33.718	26.849	43.303			117.725	
C2494	8.786	5.253	2.751 2.927					174.365	55.020	41.098					120.623	
L2495	9.394	5.070	1.700 1.679	1.580	0.948 0.921			173.839	52.151	43.883	26.833	23.201 25.115			128.098	
P2496		4.746	2.074 2.310	1.799 1.939	3.597 3.490				62.950	34.412	24.704	50.134				

7.2. Ramachandran plot of twenty lowest energy structures of VWF C3

-NMR

Ramachandran Plot
C3 (20 models)



Plot statistics

Residues in most favoured regions [A,B,L]	795	68.5%
Residues in additional allowed regions [a,b,l,p]	356	30.7%
Residues in generously allowed regions [~a,~b,~l,~p]	9	0.8%
Residues in disallowed regions	0	0.0%
	----	------
Number of non-glycine and non-proline residues	1160	100.0%
Number of end-residues (excl. Gly and Pro)	20	
Number of glycine residues (shown as triangles)	120	
Number of proline residues	60	

Total number of residues	1360	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.
Model numbers shown inside each data point.

8. References

1. Friedland G. Discovery of the function of the heart and circulation of blood. Cardiovasc J Afr. 2009;20(3):160.

2. Fastag E, Varon J, Sternbach G. Richard Lower: the origins of blood transfusion. J Emerg Med. 2013;44(6):1146-50.

3. Saito H, Matsushita T, Kojima T. Historical perspective and future direction of coagulation research. J Thromb Haemost. 2011;9 Suppl 1:352-63.

4. Rasche H. Haemostasis and thrombosis: an overview. European Heart Journal Supplements. 2001;3(suppl_Q):Q3-Q7.

5. Smith SA, Travers RJ, Morrissey JH. How it all starts: Initiation of the clotting cascade. Crit Rev Biochem Mol Biol. 2015;50(4):326-36.

6. Davie EW, Ratnoff OD. WATERFALL SEQUENCE FOR INTRINSIC BLOOD CLOTTING. Science. 1964;145(3638):1310-2.

7. Macfarlane RG. AN ENZYME CASCADE IN THE BLOOD CLOTTING MECHANISM, AND ITS FUNCTION AS A BIOCHEMICAL AMPLIFIER. Nature. 1964;202:498-9.

8. Hoffman M. A cell-based model of coagulation and the role of factor VIIa. Blood Rev. 2003;17 Suppl 1:S1-5.

9. Silverberg M, Dunn JT, Garen L, Kaplan AP. Autoactivation of human Hageman factor. Demonstration utilizing a synthetic substrate. J Biol Chem. 1980;255(15):7281-6.

10. Wu Y. Contact pathway of coagulation and inflammation. Thromb J. 2015;13:17.

11. Wiggins RC, Cochrane CC. The autoactivation of rabbit Hageman factor. J Exp Med. 1979;150(5):1122-33.

12. Grover SP, Mackman N. Intrinsic Pathway of Coagulation and Thrombosis. Arterioscler Thromb Vasc Biol. 2019;39(3):331-8.

13. Renné T, Schmaier AH, Nickel KF, Blombäck M, Maas C. In vivo roles of factor XII. Blood. 2012;120(22):4296-303.

14. Ott I. Inhibitors of the initiation of coagulation. Br J Clin Pharmacol. 2011;72(4):547-52.

15. Sallah S. Inhibitors to clotting factors. Ann Hematol. 1997;75(1-2):1-7.

16. Frebelius S, Isaksson S, Swedenborg J. Thrombin inhibition by antithrombin III on the subendothelium is explained by the isoform AT beta. Arterioscler Thromb Vasc Biol. 1996;16(10):1292-7.

17. Kreuz W. von Willebrand's disease: from discovery to therapy - milestones in the last 25 years. Haemophilia. 2008;14 Suppl 5:1-2.

18. Ikeda Y, Handa M, Kawano K, Kamata T, Murata M, Araki Y, et al. The role of von Willebrand factor and fibrinogen in platelet aggregation under varying shear stress. J Clin Invest. 1991;87(4):1234-40.
19. Sorrentino S, Studt JD, Medalia O, Tanuj Sapra K. Roll, adhere, spread and contract: structural mechanics of platelet function. Eur J Cell Biol. 2015;94(3-4):129-38.
20. Peyvandi F, Garagiola I, Baronciani L. Role of von Willebrand factor in the haemostasis. Blood Transfus. 2011;9 Suppl 2:s3-8.
21. Federici AB. The factor VIII/von Willebrand factor complex: basic and clinical issues. Haematologica. 2003;88(6):Erep02.
22. Jaffe EA, Hoyer LW, Nachman RL. Synthesis of von Willebrand factor by cultured human endothelial cells. Proc Natl Acad Sci U S A. 1974;71(5):1906-9.
23. Bowie EJ, Solberg LA, Jr., Fass DN, Johnson CM, Knutson GJ, Stewart ML, et al. Transplantation of normal bone marrow into a pig with severe von Willebrand's disease. J Clin Invest. 1986;78(1):26-30.
24. Wagner DD, Olmsted JB, Marder VJ. Immunolocalization of von Willebrand protein in Weibel-Palade bodies of human endothelial cells. J Cell Biol. 1982;95(1):355-60.
25. Sporn LA, Chavin SI, Marder VJ, Wagner DD. Biosynthesis of von Willebrand protein by human megakaryocytes. J Clin Invest. 1985;76(3):1102-6.
26. Jeanneau C, Avner P, Sultan Y. Use of monoclonal antibody and colloidal gold in E.M. localization of von Willebrand factor in megakaryocytes and platelets. Cell Biol Int Rep. 1984;8(10):841-8.
27. Kanaji S, Fahs SA, Shi Q, Haberichter SL, Montgomery RR. Contribution of platelet vs. endothelial VWF to platelet adhesion and hemostasis. J Thromb Haemost. 2012;10(8):1646-52.
28. Yamamoto K, de Waard V, Fearns C, Loskutoff DJ. Tissue distribution and regulation of murine von Willebrand factor gene expression in vivo. Blood. 1998;92(8):2791-801.
29. Rand JH, Badimon L, Gordon RE, Uson RR, Fuster V. Distribution of von Willebrand factor in porcine intima varies with blood vessel type and location. Arteriosclerosis. 1987;7(3):287-91.
30. Senis YA, Richardson M, Tinlin S, Maurice DH, Giles AR. Changes in the pattern of distribution of von Willebrand factor in rat aortic endothelial cells following thrombin generation in vivo. British Journal of Haematology. 1996;93(1):195-203.
31. Ginsburg D, Handin RI, Bonthron DT, Donlon TA, Bruns GA, Latt SA, et al. Human von Willebrand factor (vWF): isolation of complementary DNA (cDNA) clones and chromosomal localization. Science. 1985;228(4706):1401-6.

32. Mancuso DJ, Tuley EA, Westfield LA, Lester-Mancuso TL, Le Beau MM, Sorace JM, et al. Human von Willebrand factor gene and pseudogene: structural analysis and differentiation by polymerase chain reaction. Biochemistry. 1991;30(1):253-69.
33. Shelton-Inloes BB, Chehab FF, Mannucci PM, Federici AB, Sadler JE. Gene deletions correlate with the development of alloantibodies in von Willebrand disease. J Clin Invest. 1987;79(5):1459-65.
34. Sadler JE. Biochemistry and genetics of von Willebrand factor. Annu Rev Biochem. 1998;67:395-424.
35. Colombatti A, Bonaldo P. The superfamily of proteins with von Willebrand factor type A-like domains: one theme common to components of extracellular matrix, hemostasis, cellular adhesion, and defense mechanisms. Blood. 1991;77(11):2305-15.
36. Shaw SK, Cepek KL, Murphy EA, Russell GJ, Brenner MB, Parker CM. Molecular cloning of the human mucosal lymphocyte integrin alpha E subunit. Unusual structure and restricted RNA distribution. J Biol Chem. 1994;269(8):6016-25.
37. Zhou YF, Eng ET, Zhu J, Lu C, Walz T, Springer TA. Sequence and structure relationships within von Willebrand factor. Blood. 2012;120(2):449-58.
38. Titani K, Kumar S, Takio K, Ericsson LH, Wade RD, Ashida K, et al. Amino acid sequence of human von Willebrand factor. Biochemistry. 1986;25(11):3171-84.
39. Mancuso DJ, Tuley EA, Westfield LA, Worrall NK, Shelton-Inloes BB, Sorace JM, et al. Structure of the gene for human von Willebrand factor. J Biol Chem. 1989;264(33):19514-27.
40. Fahey RC, Hunt JS, Windham GC. On the cysteine and cystine content of proteins. Differences between intracellular and extracellular proteins. J Mol Evol. 1977;10(2):155-60.
41. Pannekoek H, Voorberg J. Molecular cloning, expression and assembly of multimeric von Willebrand factor. Baillieres Clin Haematol. 1989;2(4):879-96.
42. Isaacs NW. Cystine knots. Curr Opin Struct Biol. 1995;5(3):391-5.
43. Katsumi A, Tuley EA, Bodó I, Sadler JE. Localization of disulfide bonds in the cystine knot domain of human von Willebrand factor. J Biol Chem. 2000;275(33):25585-94.
44. Collado-González M, González Espinosa Y, Goycoolea FM. Interaction Between Chitosan and Mucin: Fundamentals and Applications. Biomimetics (Basel). 2019;4(2).
45. Fiebig JE, Weidauer SE, Qiu LY, Bauer M, Schmieder P, Beerbaum M, et al. The clip-segment of the von Willebrand domain 1 of the BMP modulator protein Crossveinless 2 is preformed. Molecules. 2013;18(10):11658-82.
46. Grasberger BL, Clore GM, Gronenborn AM. High-resolution structure of Ascaris trypsin inhibitor in solution: direct evidence for a pH-induced conformational transition in the reactive site. Structure. 1994;2(7):669-78.

47. Xu ER, Blythe EE, Fischer G, Hyvönen M. Structural analyses of von Willebrand factor C domains of collagen 2A and CCN3 reveal an alternative mode of binding to bone morphogenetic protein-2. J Biol Chem. 2017;292(30):12516-27.

48. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. 2015;43(Database issue):D257-60.

49. O'Leary JM, Hamilton JM, Deane CM, Valeyev NV, Sandell LJ, Downing AK. Solution structure and dynamics of a prototypical chordin-like cysteine-rich repeat (von Willebrand Factor type C module) from collagen IIA. J Biol Chem. 2004;279(51):53857-66.

50. Luo H, Nijveen H. Understanding and identifying amino acid repeats. Brief Bioinform. 2014;15(4):582-91.

51. Reininger AJ. Function of von Willebrand factor in haemostasis and thrombosis. Haemophilia. 2008;14 Suppl 5:11-26.

52. Foster PA, Fulcher CA, Marti T, Titani K, Zimmerman TS. A major factor VIII binding domain resides within the amino-terminal 272 amino acid residues of von Willebrand factor. J Biol Chem. 1987;262(18):8443-6.

53. Wu JJ, Fujikawa K, McMullen BA, Chung DW. Characterization of a core binding site for ADAMTS-13 in the A2 domain of von Willebrand factor. Proc Natl Acad Sci U S A. 2006;103(49):18470-4.

54. Wagner DD. The Weibel-Palade body: the storage granule for von Willebrand factor and P-selectin. Thromb Haemost. 1993;70(1):105-10.

55. Padilla A, Moake JL, Bernardo A, Ball C, Wang Y, Arya M, et al. P-selectin anchors newly released ultralarge von Willebrand factor multimers to the endothelial cell surface. Blood. 2004;103(6):2150-6.

56. Fretto LJ, Fowler WE, McCaslin DR, Erickson HP, McKee PA. Substructure of human von Willebrand factor. Proteolysis by V8 and characterization of two functional domains. J Biol Chem. 1986;261(33):15679-89.

57. Lenting PJ, van Mourik JA, Mertens K. The life cycle of coagulation factor VIII in view of its structure and function. Blood. 1998;92(11):3983-96.

58. Vehar GA, Keyt B, Eaton D, Rodriguez H, O'Brien DP, Rotblat F, et al. Structure of human factor VIII. Nature. 1984;312(5992):337-42.

59. Swieringa F, Kuijpers MJ, Lamers MM, van der Meijden PE, Heemskerk JW. Rate-limiting roles of the tenase complex of factors VIII and IX in platelet procoagulant activity and formation of platelet-fibrin thrombi under flow. Haematologica. 2015;100(6):748-56.

60. Sadler JE, Budde U, Eikenboom JC, Favaloro EJ, Hill FG, Holmberg L, et al. Update on the pathophysiology and classification of von Willebrand disease: a report of the Subcommittee on von Willebrand Factor. J Thromb Haemost. 2006;4(10):2103-14.
61. Federici AB, Berntorp E, Lee CA. The 80th anniversary of von Willebrand's disease: history, management and research. Haemophilia. 2006;12(6):563-72.
62. Shiltagh N, Kirkpatrick J, Cabrita LD, McKinnon TA, Thalassinos K, Tuddenham EG, et al. Solution structure of the major factor VIII binding region on von Willebrand factor. Blood. 2014;123(26):4143-51.
63. Dong X, Leksa NC, Chhabra ES, Arndt JW, Lu Q, Knockenhauer KE, et al. The von Willebrand factor D'D3 assembly and structural principles for factor VIII binding and concatemer biogenesis. Blood. 2019;133(14):1523-33.
64. Yee A, Oleskie AN, Dosey AM, Kretz CA, Gildersleeve RD, Dutta S, et al. Visualization of an N-terminal fragment of von Willebrand factor in complex with factor VIII. Blood. 2015;126(8):939-42.
65. Chiu PL, Bou-Assaf GM, Chhabra ES, Chambers MG, Peters RT, Kulman JD, et al. Mapping the interaction between factor VIII and von Willebrand factor by electron microscopy and mass spectrometry. Blood. 2015;126(8):935-8.
66. Dagil L, Troelsen KS, Bolt G, Thim L, Wu B, Zhao X, et al. Interaction Between the a3 Region of Factor VIII and the TIL'E' Domains of the von Willebrand Factor. Biophys J. 2019;117(3):479-89.
67. Saenko EL, Scandella D. The acidic region of the factor VIII light chain and the C2 domain together form the high affinity binding site for von willebrand factor. J Biol Chem. 1997;272(29):18007-14.
68. Przeradzka MA, Meems H, van der Zwaan C, Ebberink E, van den Biggelaar M, Mertens K, et al. The D' domain of von Willebrand factor requires the presence of the D3 domain for optimal factor VIII binding. Biochem J. 2018;475(17):2819-30.
69. Hilbert L, Jorieux S, Proulle V, Favier R, Goudemand J, Parquet A, et al. Two novel mutations, Q1053H and C1060R, located in the D3 domain of von Willebrand factor, are responsible for decreased FVIII-binding capacity. Br J Haematol. 2003;120(4):627-32.
70. Hilbert L, D'Oiron R, Fressinaud E, Meyer D, Mazurier C, Disease INoMAivW. First identification and expression of a type 2N von Willebrand disease mutation (E1078K) located in exon 25 of von Willebrand factor gene. J Thromb Haemost. 2. England2004. p. 2271-3.
71. Jorieux S, Gaucher C, Goudemand J, Mazurier C. A novel mutation in the D3 domain of von Willebrand factor markedly decreases its ability to bind factor VIII and affects its multimerization. Blood. 1998;92(12):4663-70.

72. Leyte A, Verbeet MP, Brodniewicz-Proba T, Van Mourik JA, Mertens K. The interaction between human blood-coagulation factor VIII and von Willebrand factor. Characterization of a high-affinity binding site on factor VIII. Biochem J. 1989;257(3):679-83.
73. Hamer RJ, Koedam JA, Beeser-Visser NH, Sixma JJ. The effect of thrombin on the complex between factor VIII and von Willebrand factor. Eur J Biochem. 1987;167(2):253-9.
74. Butenas S, van 't Veer C, Mann KG. Evaluation of the initiation phase of blood coagulation using ultrasensitive assays for serine proteases. J Biol Chem. 1997;272(34):21527-33.
75. Eaton D, Rodriguez H, Vehar GA. Proteolytic processing of human factor VIII. Correlation of specific cleavages by thrombin, factor Xa, and activated protein C with activation and inactivation of factor VIII coagulant activity. Biochemistry. 1986;25(2):505-12.
76. Neuenschwander PF, Jesty J. Thrombin-activated and factor Xa-activated human factor VIII: differences in cofactor activity and decay rate. Arch Biochem Biophys. 1992;296(2):426-34.
77. Koedam JA, Hamer RJ, Beeser-Visser NH, Bouma BN, Sixma JJ. The effect of von Willebrand factor on activation of factor VIII by factor Xa. Eur J Biochem. 1990;189(2):229-34.
78. Orfeo T, Elsman R, Gissel M, Mann KG, Butenas S. Activation, activity and inactivation of factor VIII in factor VIII products. Haemophilia. 2016;22(3):462-73.
79. Flood VH, Slobodianuk TL, Keesler D, Lohmeier HK, Fahs S, Zhang L, et al. von Willebrand factor binding to myosin assists in coagulation. Blood Adv. 2020;4(1):174-80.
80. Cruz MA, Yuan H, Lee JR, Wise RJ, Handin RI. Interaction of the von Willebrand factor (vWF) with collagen. Localization of the primary collagen-binding site by analysis of recombinant vWF a domain polypeptides. J Biol Chem. 1995;270(18):10822-7.
81. Bienkowska J, Cruz M, Atiemo A, Handin R, Liddington R. The von willebrand factor A3 domain does not contain a metal ion-dependent adhesion site motif. J Biol Chem. 1997;272(40):25162-7.
82. Wu D, Vanhoorelbeke K, Cauwenberghs N, Meiring M, Depraetere H, Kotze HF, et al. Inhibition of the von Willebrand (VWF)-collagen interaction by an antihuman VWF monoclonal antibody results in abolition of in vivo arterial platelet thrombus formation in baboons. Blood. 2002;99(10):3623-8.
83. Lee JO, Rieu P, Arnaout MA, Liddington R. Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). Cell. 1995;80(4):631-8.
84. Qu A, Leahy DJ. Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, alpha L beta 2) integrin. Proc Natl Acad Sci U S A. 1995;92(22):10277-81.

85. Tuckwell D. Evolution of von Willebrand factor A (VWA) domains. Biochem Soc Trans. 1999;27(6):835-40.
86. Loftus JC, Liddington RC. Cell adhesion in vascular biology. New insights into integrin-ligand interaction. J Clin Invest. 1997;99(10):2302-6.
87. Whittaker CA, Hynes RO. Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. Mol Biol Cell. 2002;13(10):3369-87.
88. Emsley J, Cruz M, Handin R, Liddington R. Crystal structure of the von Willebrand Factor A1 domain and implications for the binding of platelet glycoprotein Ib. J Biol Chem. 1998;273(17):10396-401.
89. Huizinga EG, Martijn van der Plas R, Kroon J, Sixma JJ, Gros P. Crystal structure of the A3 domain of human von Willebrand factor: implications for collagen binding. Structure. 1997;5(9):1147-56.
90. van der Plas RM, Gomes L, Marquart JA, Vink T, Meijers JC, de Groot PG, et al. Binding of von Willebrand factor to collagen type III: role of specific amino acids in the collagen binding domain of vWF and effects of neighboring domains. Thromb Haemost. 2000;84(6):1005-11.
91. Romijn RA, Bouma B, Wuyster W, Gros P, Kroon J, Sixma JJ, et al. Identification of the collagen-binding site of the von Willebrand factor A3-domain. J Biol Chem. 2001;276(13):9985-91.
92. Romijn RA, Westein E, Bouma B, Schiphorst ME, Sixma JJ, Lenting PJ, et al. Mapping the collagen-binding site in the von Willebrand factor-A3 domain. J Biol Chem. 2003;278(17):15035-9.
93. Nishida N, Sumikawa H, Sakakura M, Shimba N, Takahashi H, Terasawa H, et al. Collagen-binding mode of vWF-A3 domain determined by a transferred cross-saturation experiment. Nat Struct Biol. 2003;10(1):53-8.
94. Santoro SA. Preferential binding of high molecular weight forms of von Willebrand factor to fibrillar collagen. Biochim Biophys Acta. 1983;756(1):123-6.
95. Shoulders MD, Raines RT. Collagen structure and stability. Annu Rev Biochem. 2009;78:929-58.
96. Kadler KE, Baldock C, Bella J, Boot-Handford RP. Collagens at a glance. J Cell Sci. 2007;120(Pt 12):1955-8.
97. Brondijk TH, Bihan D, Farndale RW, Huizinga EG. Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3 domain complex. Proc Natl Acad Sci U S A. 2012;109(14):5253-8.

98. Hoylaerts MF, Yamamoto H, Nuyts K, Vreys I, Deckmyn H, Vermylen J. von Willebrand factor binds to native collagen VI primarily via its A1 domain. Biochem J. 1997;324 (Pt 1)(Pt 1):185-91.

99. Ruggeri ZM. Platelet adhesion under flow. Microcirculation. 2009;16(1):58-83.

100. Lisman T, Raynal N, Groeneveld D, Maddox B, Peachey AR, Huizinga EG, et al. A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. Blood. 2006;108(12):3753-6.

101. Lankhof H, van Hoeij M, Schiphorst ME, Bracke M, Wu YP, Ijsseldijk MJ, et al. A3 domain is essential for interaction of von Willebrand factor with collagen type III. Thromb Haemost. 1996;75(6):950-8.

102. Flood VH, Schlauderaff AC, Haberichter SL, Slobodianuk TL, Jacobi PM, Bellissimo DB, et al. Crucial role for the VWF A1 domain in binding to type IV collagen. Blood. 2015;125(14):2297-304.

103. Slobodianuk TL, Kochelek C, Foeckler J, Kalloway S, Weiler H, Flood VH. Defective collagen binding and increased bleeding in a murine model of von Willebrand disease affecting collagen IV binding. J Thromb Haemost. 2019;17(1):63-71.

104. Bonnefoy A, Romijn RA, Vandervoort PA, I VANR, Vermylen J, Hoylaerts MF. von Willebrand factor A1 domain can adequately substitute for A3 domain in recruitment of flowing platelets to collagen. J Thromb Haemost. 2006;4(10):2151-61.

105. Morales LD, Martin C, Cruz MA. The interaction of von Willebrand factor-A1 domain with collagen: mutation G1324S (type 2M von Willebrand disease) impairs the conformational change in A1 domain induced by collagen. J Thromb Haemost. 2006;4(2):417-25.

106. Watson SP, Gibbins J. Collagen receptor signalling in platelets: extending the role of the ITAM. Immunol Today. 1998;19(6):260-4.

107. Savage B, Saldívar E, Ruggeri ZM. Initiation of platelet adhesion by arrest onto fibrinogen or translocation on von Willebrand factor. Cell. 1996;84(2):289-97.

108. Weisel JW, Nagaswami C, Vilaire G, Bennett JS. Examination of the platelet membrane glycoprotein IIb-IIIa complex and its interaction with fibrinogen and other ligands by electron microscopy. J Biol Chem. 1992;267(23):16637-43.

109. Andrews RK, López JA, Berndt MC. Molecular mechanisms of platelet adhesion and activation. Int J Biochem Cell Biol. 1997;29(1):91-105.

110. Uff S, Clemetson JM, Harrison T, Clemetson KJ, Emsley J. Crystal structure of the platelet glycoprotein Ib(alpha) N-terminal domain reveals an unmasking mechanism for receptor activation. J Biol Chem. 2002;277(38):35657-63.

111. Huizinga EG, Tsuji S, Romijn RA, Schiphorst ME, de Groot PG, Sixma JJ, et al. Structures of glycoprotein Ibalpha and its complex with von Willebrand factor A1 domain. Science. 2002;297(5584):1176-9.

112. Dumas JJ, Kumar R, McDonagh T, Sullivan F, Stahl ML, Somers WS, et al. Crystal structure of the wild-type von Willebrand factor A1-glycoprotein Ibalpha complex reveals conformation differences with a complex bearing von Willebrand disease mutations. J Biol Chem. 2004;279(22):23327-34.

113. Reininger AJ, Heijnen HF, Schumann H, Specht HM, Schramm W, Ruggeri ZM. Mechanism of platelet adhesion to von Willebrand factor and microparticle formation under high shear stress. Blood. 2006;107(9):3537-45.

114. Tischer A, Cruz MA, Auton M. The linker between the D3 and A1 domains of vWF suppresses A1-GPIbα catch bonds by site-specific binding to the A1 domain. Protein Sci. 2013;22(8):1049-59.

115. Shattil SJ, Newman PJ. Integrins: dynamic scaffolds for adhesion and signaling in platelets. Blood. 2004;104(6):1606-15.

116. Durrant TN, van den Bosch MT, Hers I. Integrin α(IIb)β(3) outside-in signaling. Blood. 2017;130(14):1607-19.

117. Jennings LK, Phillips DR. Purification of glycoproteins IIb and III from human platelet plasma membranes and characterization of a calcium-dependent glycoprotein IIb-III complex. J Biol Chem. 1982;257(17):10458-66.

118. Shattil SJ, Kashiwagi H, Pampori N. Integrin signaling: the platelet paradigm. Blood. 1998;91(8):2645-57.

119. Takagi J, Petre BM, Walz T, Springer TA. Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. Cell. 2002;110(5):599-11.

120. Faull RJ, Ginsberg MH. Inside-out signaling through integrins. J Am Soc Nephrol. 1996;7(8):1091-7.

121. Shattil SJ, Kim C, Ginsberg MH. The final steps of integrin activation: the end game. Nat Rev Mol Cell Biol. 2010;11(4):288-300.

122. French DL, Seligsohn U. Platelet glycoprotein IIb/IIIa receptors and Glanzmann's thrombasthenia. Arterioscler Thromb Vasc Biol. 2000;20(3):607-10.

123. Li Z, Delaney MK, O'Brien KA, Du X. Signaling during platelet adhesion and activation. Arterioscler Thromb Vasc Biol. 2010;30(12):2341-9.

124. Aslan JE, Itakura A, Gertz JM, McCarty OJ. Platelet shape change and spreading. Methods Mol Biol. 2012;788:91-100.

125. Xu ER, von Bulow S, Chen PC, Lenting PJ, Kolsek K, Aponte-Santamaria C, et al. Structure and dynamics of the platelet integrin-binding C4 domain of von Willebrand factor. Blood. 2019;133(4):366-76.
126. Springer TA. Biology and physics of von Willebrand factor concatamers. J Thromb Haemost. 2011;9 Suppl 1(0 1):130-43.
127. Naimushin YA, Mazurov AV. Von Willebrand factor can support platelet aggregation via interaction with activated GPIIb-IIIa and GPIb. Platelets. 2004;15(7):419-25.
128. Joo SJ. Mechanisms of Platelet Activation and Integrin αIIβ3. Korean Circ J. 2012;42(5):295-301.
129. Zhang JL, Qiu LY, Kotzsch A, Weidauer S, Patterson L, Hammerschmidt M, et al. Crystal structure analysis reveals how the Chordin family member crossveinless 2 blocks BMP-2 receptor binding. Dev Cell. 2008;14(5):739-50.
130. Marti T, Rosselet SJ, Titani K, Walsh KA. Identification of disulfide-bridged substructures within human von Willebrand factor. Biochemistry. 1987;26(25):8099-109.
131. Hatahet F, Ruddock LW. Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. Antioxid Redox Signal. 2009;11(11):2807-50.
132. Sadler JE. von Willebrand factor assembly and secretion. J Thromb Haemost. 2009;7 Suppl 1:24-7.
133. Araki K, Nagata K. Protein folding and quality control in the ER. Cold Spring Harb Perspect Biol. 2011;3(11):a007526.
134. Canis K, McKinnon TA, Nowak A, Haslam SM, Panico M, Morris HR, et al. Mapping the N-glycome of human von Willebrand factor. Biochem J. 2012;447(2):217-28.
135. Wagner DD, Mayadas T, Marder VJ. Initial glycosylation and acidic pH in the Golgi apparatus are required for multimerization of von Willebrand factor. J Cell Biol. 1986;102(4):1320-4.
136. Debeire P, Montreuil J, Samor B, Mazurier C, Goudemand M, van Halbeek H, et al. Structure determination of the major asparagine-linked sugar chain of human factor VIII--von Willebrand factor. FEBS Lett. 1983;151(1):22-6.
137. Matsui T, Titani K, Mizuochi T. Structures of the asparagine-linked oligosaccharide chains of human von Willebrand factor. Occurrence of blood group A, B, and H(O) structures. J Biol Chem. 1992;267(13):8723-31.
138. Kano T, Kondo K, Hamako J, Matsushita F, Sakai K, Matsui T. Effects of plasma glycosyltransferase on the ABO(H) blood group antigens of human von Willebrand factor. Int J Hematol. 2018;108(2):139-44.

139. McGrath RT, McKinnon TA, Byrne B, O'Kennedy R, Terraube V, McRae E, et al. Expression of terminal alpha2-6-linked sialic acid on von Willebrand factor specifically enhances proteolysis by ADAMTS13. Blood. 2010;115(13):2666-73.
140. Ward S, O'Sullivan JM, O'Donnell JS. von Willebrand factor sialylation-A critical regulator of biological function. J Thromb Haemost. 2019;17(7):1018-29.
141. McKinnon TA, Chion AC, Millington AJ, Lane DA, Laffan MA. N-linked glycosylation of VWF modulates its interaction with ADAMTS13. Blood. 2008;111(6):3042-9.
142. Sodetz JM, Pizzo SV, McKee PA. Relationship of sialic acid to function and in vivo survival of human factor VIII/von Willebrand factor protein. J Biol Chem. 1977;252(15):5538-46.
143. Franchini M, Capra F, Targher G, Montagnana M, Lippi G. Relationship between ABO blood group and von Willebrand factor levels: from biology to clinical implications. Thromb J. 2007;5:14.
144. Murray GP, Post SR, Post GR. ABO blood group is a determinant of von Willebrand factor protein levels in human pulmonary endothelial cells. J Clin Pathol. 2020;73(6):347-9.
145. Gill JC, Endres-Brooks J, Bauer PJ, Marks WJ, Jr., Montgomery RR. The effect of ABO blood group on the diagnosis of von Willebrand disease. Blood. 1987;69(6):1691-5.
146. Mohlke KL, Purkayastha AA, Westrick RJ, Smith PL, Petryniak B, Lowe JB, et al. Mvwf, a dominant modifier of murine von Willebrand factor, results from altered lineage-specific expression of a glycosyltransferase. Cell. 1999;96(1):111-20.
147. Bowen DJ. An influence of ABO blood group on the rate of proteolysis of von Willebrand factor by ADAMTS13. J Thromb Haemost. 2003;1(1):33-40.
148. Haberichter SL, Fahs SA, Montgomery RR. von Willebrand factor storage and multimerization: 2 independent intracellular processes. Blood. 2000;96(5):1808-15.
149. Stockschlaeder M, Schneppenheim R, Budde U. Update on von Willebrand factor multimers: focus on high-molecular-weight multimers and their role in hemostasis. Blood Coagul Fibrinolysis. 2014;25(3):206-16.
150. Hassan MI, Saxena A, Ahmad F. Structure and function of von Willebrand factor. Blood Coagul Fibrinolysis. 2012;23(1):11-22.
151. Zhou YF, Springer TA. Highly reinforced structure of a C-terminal dimerization domain in von Willebrand factor. Blood. 2014;123(12):1785-93.
152. Lippok S, Kolšek K, Löf A, Eggert D, Vanderlinden W, Müller JP, et al. von Willebrand factor is dimerized by protein disulfide isomerase. Blood. 2016;127(9):1183-91.

153. McKinnon TA, Goode EC, Birdsey GM, Nowak AA, Chan AC, Lane DA, et al. Specific N-linked glycosylation sites modulate synthesis and secretion of von Willebrand factor. Blood. 2010;116(4):640-8.
154. Lenting PJ, Christophe OD, Denis CV. von Willebrand factor biosynthesis, secretion, and clearance: connecting the far ends. Blood. 2015;125(13):2019-28.
155. Carew JA, Browning PJ, Lynch DC. Sulfation of von Willebrand factor. Blood. 1990;76(12):2530-9.
156. Canis K, McKinnon TA, Nowak A, Panico M, Morris HR, Laffan M, et al. The plasma von Willebrand factor O-glycome comprises a surprising variety of structures including ABH antigens and disialosyl motifs. J Thromb Haemost. 2010;8(1):137-45.
157. Gashash EA, Aloor A, Li D, Zhu H, Xu XQ, Xiao C, et al. An Insight into Glyco-Microheterogeneity of Plasma von Willebrand Factor by Mass Spectrometry. J Proteome Res. 2017;16(9):3348-62.
158. Samor B, Michalski JC, Mazurier C, Goudemand M, De Waard P, Vliegenthart JF, et al. Primary structure of the major O-glycosidically linked carbohydrate unit of human von Willebrand factor. Glycoconj J. 1989;6(3):263-70.
159. Solecka BA, Weise C, Laffan MA, Kannicht C. Site-specific analysis of von Willebrand factor O-glycosylation. J Thromb Haemost. 2016;14(4):733-46.
160. Schulte am Esch J, 2nd, Robson SC, Knoefel WT, Eisenberger CF, Peiper M, Rogiers X. Impact of O-linked glycosylation of the VWF-A1-domain flanking regions on platelet interaction. Br J Haematol. 2005;128(1):82-90.
161. Badirou I, Kurdi M, Legendre P, Rayes J, Bryckaert M, Casari C, et al. In vivo analysis of the role of O-glycosylations of von Willebrand factor. PLoS One. 2012;7(5):e37508.
162. Nowak AA, Canis K, Riddell A, Laffan MA, McKinnon TA. O-linked glycosylation of von Willebrand factor modulates the interaction with platelet receptor glycoprotein Ib under static and shear stress conditions. Blood. 2012;120(1):214-22.
163. van Schooten CJ, Denis CV, Lisman T, Eikenboom JC, Leebeek FW, Goudemand J, et al. Variations in glycosylation of von Willebrand factor with O-linked sialylated T antigen are associated with its plasma levels. Blood. 2007;109(6):2430-7.
164. Paroutis P, Touret N, Grinstein S. The pH of the secretory pathway: measurement, determinants, and regulation. Physiology (Bethesda). 2004;19:207-15.
165. Mayadas TN, Wagner DD. In vitro multimerization of von Willebrand factor is triggered by low pH. Importance of the propolypeptide and free sulfhydryls. J Biol Chem. 1989;264(23):13497-503.

166. Thomas G. Furin at the cutting edge: from protein traffic to embryogenesis and disease. Nat Rev Mol Cell Biol. 2002;3(10):753-66.
167. Elferich J, Williamson DM, David LL, Shinde U. Determination of Histidine pKa Values in the Propeptides of Furin and Proprotein Convertase 1/3 Using Histidine Hydrogen-Deuterium Exchange Mass Spectrometry. Anal Chem. 2015;87(15):7909-17.
168. Feliciangeli SF, Thomas L, Scott GK, Subbian E, Hung CH, Molloy SS, et al. Identification of a pH sensor in the furin propeptide that regulates enzyme activation. J Biol Chem. 2006;281(23):16108-16.
169. Purvis AR, Sadler JE. A covalent oxidoreductase intermediate in propeptide-dependent von Willebrand factor multimerization. J Biol Chem. 2004;279(48):49982-8.
170. Dang LT, Purvis AR, Huang RH, Westfield LA, Sadler JE. Phylogenetic and functional analysis of histidine residues essential for pH-dependent multimerization of von Willebrand factor. J Biol Chem. 2011;286(29):25763-9.
171. Huang RH, Wang Y, Roth R, Yu X, Purvis AR, Heuser JE, et al. Assembly of Weibel-Palade body-like tubules from N-terminal domains of von Willebrand factor. Proc Natl Acad Sci U S A. 2008;105(2):482-7.
172. Mayadas TN, Wagner DD. Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. Proc Natl Acad Sci U S A. 1992;89(8):3531-5.
173. Wise RJ, Pittman DD, Handin RI, Kaufman RJ, Orkin SH. The propeptide of von Willebrand factor independently mediates the assembly of von Willebrand multimers. Cell. 1988;52(2):229-36.
174. Verweij CL, Hart M, Pannekoek H. Proteolytic cleavage of the precursor of von Willebrand factor is not essential for multimer formation. J Biol Chem. 1988;263(17):7921-4.
175. Hilbert L, Nurden P, Caron C, Nurden AT, Goudemand J, Meyer D, et al. Type 2N von Willebrand disease due to compound heterozygosity for R854Q and a novel R763G mutation at the cleavage site of von Willebrand factor propeptide. Thromb Haemost. 2006;96(3):290-4.
176. Dong Z, Thoma RS, Crimmins DL, McCourt DW, Tuley EA, Sadler JE. Disulfide bonds required to assemble functional von Willebrand factor multimers. J Biol Chem. 1994;269(9):6753-8.
177. Purvis AR, Gross J, Dang LT, Huang RH, Kapadia M, Townsend RR, et al. Two Cys residues essential for von Willebrand factor multimer assembly in the Golgi. Proc Natl Acad Sci U S A. 2007;104(40):15647-52.
178. Xiang Y, Hwa J. Regulation of VWF expression, and secretion in health and disease. Curr Opin Hematol. 2016;23(3):288-93.

179. Williams SB, McKeown LP, Krutzsch H, Hansmann K, Gralnick HR. Purification and characterization of human platelet von Willebrand factor. Br J Haematol. 1994;88(3):582-91.
180. Ruggeri ZM. Structure and function of von Willebrand factor. Thromb Haemost. 1999;82(2):576-84.
181. Rand JH, Glanville RW, Wu XX, Ross JM, Zangari M, Gordon RE, et al. The significance of subendothelial von Willebrand factor. Thromb Haemost. 1997;78(1):445-50.
182. Dong JF, Moake JL, Nolasco L, Bernardo A, Arceneaux W, Shrimpton CN, et al. ADAMTS-13 rapidly cleaves newly secreted ultralarge von Willebrand factor multimers on the endothelial surface under flowing conditions. Blood. 2002;100(12):4033-9.
183. Weibel ER, Palade GE. NEW CYTOPLASMIC COMPONENTS IN ARTERIAL ENDOTHELIA. J Cell Biol. 1964;23(1):101-12.
184. Nagy EE, Varga-Fekete T, Puskas A, Kelemen P, Brassai Z, Szekeres-Csiki K, et al. High circulating osteoprotegerin levels are associated with non-zero blood groups. BMC Cardiovasc Disord. 2016;16(1):106.
185. Saint-Lu N, Oortwijn BD, Pegon JN, Odouard S, Christophe OD, de Groot PG, et al. Identification of galectin-1 and galectin-3 as novel partners for von Willebrand factor. Arterioscler Thromb Vasc Biol. 2012;32(4):894-901.
186. Metcalf DJ, Nightingale TD, Zenner HL, Lui-Roberts WW, Cutler DF. Formation and function of Weibel-Palade bodies. J Cell Sci. 2008;121(Pt 1):19-27.
187. Rondaij MG, Bierings R, Kragt A, van Mourik JA, Voorberg J. Dynamics and plasticity of Weibel-Palade bodies in endothelial cells. Arterioscler Thromb Vasc Biol. 2006;26(5):1002-7.
188. Valentijn KM, Sadler JE, Valentijn JA, Voorberg J, Eikenboom J. Functional architecture of Weibel-Palade bodies. Blood. 2011;117(19):5033-43.
189. McCormack JJ, Lopes da Silva M, Ferraro F, Patella F, Cutler DF. Weibel-Palade bodies at a glance. J Cell Sci. 2017;130(21):3611-7.
190. Springer TA. von Willebrand factor, Jedi knight of the bloodstream. Blood. 2014;124(9):1412-25.
191. Zhou YF, Eng ET, Nishida N, Lu C, Walz T, Springer TA. A pH-regulated dimeric bouquet in the structure of von Willebrand factor. Embo j. 2011;30(19):4098-111.
192. Gerke V. Von Willebrand factor folds into a bouquet. Embo j. 2011;30(19):3880-1.
193. Michaux G, Abbitt KB, Collinson LM, Haberichter SL, Norman KE, Cutler DF. The physiological function of von Willebrand's factor depends on its tubular storage in endothelial Weibel-Palade bodies. Dev Cell. 2006;10(2):223-32.

194. Lui-Roberts WW, Ferraro F, Nightingale TD, Cutler DF. Aftiphilin and gamma-synergin are required for secretagogue sensitivity of Weibel-Palade bodies in endothelial cells. Mol Biol Cell. 2008;19(12):5072-81.
195. Berriman JA, Li S, Hewlett LJ, Wasilewski S, Kiskin FN, Carter T, et al. Structural organization of Weibel-Palade bodies revealed by cryo-EM of vitrified endothelial cells. Proc Natl Acad Sci U S A. 2009;106(41):17407-12.
196. Zenner HL, Collinson LM, Michaux G, Cutler DF. High-pressure freezing provides insights into Weibel-Palade body biogenesis. J Cell Sci. 2007;120(Pt 12):2117-25.
197. Erent M, Meli A, Moisoi N, Babich V, Hannah MJ, Skehel P, et al. Rate, extent and concentration dependence of histamine-evoked Weibel-Palade body exocytosis determined from individual fusion events in human endothelial cells. J Physiol. 2007;583(Pt 1):195-212.
198. Fowler WE, Fretto LJ, Hamilton KK, Erickson HP, McKee PA. Substructure of human von Willebrand factor. J Clin Invest. 1985;76(4):1491-500.
199. Nightingale TD, Pattni K, Hume AN, Seabra MC, Cutler DF. Rab27a and MyRIP regulate the amount and multimeric state of VWF released from endothelial cells. Blood. 2009;113(20):5010-8.
200. Knop M, Aareskjold E, Bode G, Gerke V. Rab3D and annexin A2 play a role in regulated secretion of vWF, but not tPA, from endothelial cells. Embo j. 2004;23(15):2982-92.
201. Bierings R, Hellen N, Kiskin N, Knipe L, Fonseca AV, Patel B, et al. The interplay between the Rab27A effectors Slp4-a and MyRIP controls hormone-evoked Weibel-Palade body exocytosis. Blood. 2012;120(13):2757-67.
202. Rojo Pulido I, Nightingale TD, Darchen F, Seabra MC, Cutler DF, Gerke V. Myosin Va acts in concert with Rab27a and MyRIP to regulate acute von-Willebrand factor release from endothelial cells. Traffic. 2011;12(10):1371-82.
203. Fukuda M. Slp4-a/granuphilin-a inhibits dense-core vesicle exocytosis through interaction with the GDP-bound form of Rab27A in PC12 cells. J Biol Chem. 2003;278(17):15390-6.
204. El-Amraoui A, Schonn JS, Küssel-Andermann P, Blanchard S, Desnos C, Henry JP, et al. MyRIP, a novel Rab effector, enables myosin VIIa recruitment to retinal melanosomes. EMBO Rep. 2002;3(5):463-70.
205. Valentijn KM, Valentijn JA, Jansen KA, Koster AJ. A new look at Weibel-Palade body structure in endothelial cells using electron tomography. J Struct Biol. 2008;161(3):447-58.
206. Turner NA, Moake JL. Factor VIII Is Synthesized in Human Endothelial Cells, Packaged in Weibel-Palade Bodies and Secreted Bound to ULVWF Strings. PLoS One. 2015;10(10):e0140740.

207. Michaux G, Pullen TJ, Haberichter SL, Cutler DF. P-selectin binds to the D'-D3 domains of von Willebrand factor in Weibel-Palade bodies. Blood. 2006;107(10):3922-4.
208. Patel KN, Soubra SH, Bellera RV, Dong JF, McMullen CA, Burns AR, et al. Differential role of von Willebrand factor and P-selectin on microvascular thrombosis in endotoxemia. Arterioscler Thromb Vasc Biol. 2008;28(12):2225-30.
209. Saidenberg-Kermanac'h N, Cohen-Solal M, Bessis N, De Vernejoul MC, Boissier MC. Role for osteoprotegerin in rheumatoid inflammation. Joint Bone Spine. 2004;71(1):9-13.
210. Zannettino AC, Holding CA, Diamond P, Atkins GJ, Kostakis P, Farrugia A, et al. Osteoprotegerin (OPG) is localized to the Weibel-Palade bodies of human vascular endothelial cells and is physically associated with von Willebrand factor. J Cell Physiol. 2005;204(2):714-23.
211. Bickel M. The role of interleukin-8 in inflammation and mechanisms of regulation. J Periodontol. 1993;64(5 Suppl):456-60.
212. Utgaard JO, Jahnsen FL, Bakka A, Brandtzaeg P, Haraldsen G. Rapid secretion of prestored interleukin 8 from Weibel-Palade bodies of microvascular endothelial cells. J Exp Med. 1998;188(9):1751-6.
213. Fiedler U, Scharpfenecker M, Koidl S, Hegen A, Grunow V, Schmidt JM, et al. The Tie-2 ligand angiopoietin-2 is stored in and rapidly released upon stimulation from endothelial cell Weibel-Palade bodies. Blood. 2004;103(11):4150-6.
214. Scholz A, Plate KH, Reiss Y. Angiopoietin-2: a multifaceted cytokine that functions in both angiogenesis and inflammation. Ann N Y Acad Sci. 2015;1347:45-51.
215. Harrison-Lavoie KJ, Michaux G, Hewlett L, Kaur J, Hannah MJ, Lui-Roberts WW, et al. P-selectin and CD63 use different mechanisms for delivery to Weibel-Palade bodies. Traffic. 2006;7(6):647-62.
216. Schnyder-Candrian S, Borsig L, Moser R, Berger EG. Localization of alpha 1,3-fucosyltransferase VI in Weibel-Palade bodies of human endothelial cells. Proc Natl Acad Sci U S A. 2000;97(15):8369-74.
217. Øynebråten I, Bakke O, Brandtzaeg P, Johansen FE, Haraldsen G. Rapid chemokine secretion from endothelial cells originates from 2 distinct compartments. Blood. 2004;104(2):314-20.
218. Ozaka T, Doi Y, Kayashima K, Fujimoto S. Weibel-Palade bodies as a storage site of calcitonin gene-related peptide and endothelin-1 in blood vessels of the rat carotid body. Anat Rec. 1997;247(3):388-94.
219. Huber D, Cramer EM, Kaufmann JE, Meda P, Massé JM, Kruithof EK, et al. Tissue-type plasminogen activator (t-PA) is stored in Weibel-Palade bodies in human endothelial cells both in vitro and in vivo. Blood. 2002;99(10):3637-45.

220. Shahbazi S, Lenting PJ, Fribourg C, Terraube V, Denis CV, Christophe OD. Characterization of the interaction between von Willebrand factor and osteoprotegerin. J Thromb Haemost. 2007;5(9):1956-62.
221. Knipe L, Meli A, Hewlett L, Bierings R, Dempster J, Skehel P, et al. A revised model for the secretion of tPA and cytokines from cultured endothelial cells. Blood. 2010;116(12):2183-91.
222. Ferraro F, Kriston-Vizi J, Metcalf DJ, Martin-Martin B, Freeman J, Burden JJ, et al. A two-tier Golgi-based control of organelle size underpins the functional plasticity of endothelial cells. Dev Cell. 2014;29(3):292-304.
223. Datta YH, Ewenstein BM. Regulated secretion in endothelial cells: biology and clinical implications. Thromb Haemost. 2001;86(5):1148-55.
224. van Mourik JA, Romani de Wit T, Voorberg J. Biogenesis and exocytosis of Weibel-Palade bodies. Histochem Cell Biol. 2002;117(2):113-22.
225. Kaufmann JE, Vischer UM. Cellular mechanisms of the hemostatic effects of desmopressin (DDAVP). J Thromb Haemost. 2003;1(4):682-9.
226. Hamilton KK, Sims PJ. Changes in cytosolic Ca2+ associated with von Willebrand factor release in human endothelial cells exposed to histamine. Study of microcarrier cell monolayers using the fluorescent probe indo-1. J Clin Invest. 1987;79(2):600-8.
227. Vischer UM, Barth H, Wollheim CB. Regulated von Willebrand factor secretion is associated with agonist-specific patterns of cytoskeletal remodeling in cultured endothelial cells. Arterioscler Thromb Vasc Biol. 2000;20(3):883-91.
228. van den Eijnden-Schrauwen Y, Atsma DE, Lupu F, de Vries RE, Kooistra T, Emeis JJ. Involvement of calcium and G proteins in the acute release of tissue-type plasminogen activator and von Willebrand factor from cultured human endothelial cells. Arterioscler Thromb Vasc Biol. 1997;17(10):2177-87.
229. van Nieuw Amerongen GP, van Delft S, Vermeer MA, Collard JG, van Hinsbergh VW. Activation of RhoA by thrombin in endothelial hyperpermeability: role of Rho kinase and protein tyrosine kinases. Circ Res. 2000;87(4):335-40.
230. Kugelmann D, Rotkopf LT, Radeva MY, Garcia-Ponce A, Walter E, Waschke J. Histamine causes endothelial barrier disruption via Ca(2+)-mediated RhoA activation and tension at adherens junctions. Sci Rep. 2018;8(1):13229.
231. Trani M, Dejana E. New insights in the control of vascular permeability: vascular endothelial-cadherin and other players. Curr Opin Hematol. 2015;22(3):267-72.
232. Michel CC, Curry FE. Microvascular permeability. Physiol Rev. 1999;79(3):703-61.

233. Kooistra MR, Corada M, Dejana E, Bos JL. Epac1 regulates integrity of endothelial cell junctions through VE-cadherin. FEBS Lett. 2005;579(22):4966-72.
234. Romani de Wit T, Rondaij MG, Hordijk PL, Voorberg J, van Mourik JA. Real-time imaging of the dynamics and secretory behavior of Weibel-Palade bodies. Arterioscler Thromb Vasc Biol. 2003;23(5):755-61.
235. André P, Denis CV, Ware J, Saffaripour S, Hynes RO, Ruggeri ZM, et al. Platelets adhere to and translocate on von Willebrand factor presented by endothelium in stimulated veins. Blood. 2000;96(10):3322-8.
236. Giblin JP, Hewlett LJ, Hannah MJ. Basal secretion of von Willebrand factor from human endothelial cells. Blood. 2008;112(4):957-64.
237. Huang J, Roth R, Heuser JE, Sadler JE. Integrin alpha(v)beta(3) on human endothelial cells binds von Willebrand factor strings under fluid shear stress. Blood. 2009;113(7):1589-97.
238. Jin SY, Skipwith CG, Shang D, Zheng XL. von Willebrand factor cleaved from endothelial cells by ADAMTS13 remains ultralarge in size. J Thromb Haemost. 72009. p. 1749-52.
239. Kawecki C, Lenting PJ, Denis CV. von Willebrand factor and inflammation. J Thromb Haemost. 2017;15(7):1285-94.
240. Siedlecki CA, Lestini BJ, Kottke-Marchant KK, Eppell SJ, Wilson DL, Marchant RE. Shear-dependent changes in the three-dimensional structure of human von Willebrand factor. Blood. 1996;88(8):2939-50.
241. Mourik MJ, Valentijn JA, Voorberg J, Koster AJ, Valentijn KM, Eikenboom J. von Willebrand factor remodeling during exocytosis from vascular endothelial cells. J Thromb Haemost. 2013;11(11):2009-19.
242. Slayter H, Loscalzo J, Bockenstedt P, Handin RI. Native conformation of human von Willebrand protein. Analysis by electron microscopy and quasi-elastic light scattering. J Biol Chem. 1985;260(14):8559-63.
243. Gragnano F, Sperlongano S, Golia E, Natale F, Bianchi R, Crisci M, et al. The Role of von Willebrand Factor in Vascular Inflammation: From Pathogenesis to Targeted Therapy. Mediators Inflamm. 2017;2017:5620314.
244. Babich V, Meli A, Knipe L, Dempster JE, Skehel P, Hannah MJ, et al. Selective release of molecules from Weibel-Palade bodies during a lingering kiss. Blood. 2008;111(11):5282-90.
245. Nightingale T, Cutler D. The secretion of von Willebrand factor from endothelial cells; an increasingly complicated story. J Thromb Haemost. 2013;11 Suppl 1:192-201.
246. Johnsen J, Lopez JA. VWF secretion: what's in a name? Blood. 2008;112(4):926-7.

247. Gralnick HR, Williams SB, McKeown LP, Krizek DM, Shafer BC, Rick ME. Platelet von Willebrand factor: comparison with plasma von Willebrand factor. Thromb Res. 1985;38(6):623-33.
248. Blair P, Flaumenhaft R. Platelet alpha-granules: basic biology and clinical correlates. Blood Rev. 2009;23(4):177-89.
249. De Meyer SF, Vandeputte N, Pareyn I, Petrus I, Lenting PJ, Chuah MK, et al. Restoration of plasma von Willebrand factor deficiency is sufficient to correct thrombus formation after gene therapy for severe von Willebrand disease. Arterioscler Thromb Vasc Biol. 2008;28(9):1621-6.
250. Harrison P, Cramer EM. Platelet alpha-granules. Blood Rev. 1993;7(1):52-62.
251. Johnston GI, Cook RG, McEver RP. Cloning of GMP-140, a granule membrane protein of platelets and endothelium: sequence similarity to proteins involved in cell adhesion and inflammation. Cell. 1989;56(6):1033-44.
252. Italiano JE, Jr., Richardson JL, Patel-Hett S, Battinelli E, Zaslavsky A, Short S, et al. Angiogenesis is regulated by a novel mechanism: pro- and antiangiogenic proteins are organized into separate platelet alpha granules and differentially released. Blood. 2008;111(3):1227-33.
253. Richardson JL, Shivdasani RA, Boers C, Hartwig JH, Italiano JE, Jr. Mechanisms of organelle transport and capture along proplatelets during platelet production. Blood. 2005;106(13):4066-75.
254. Yun SH, Sim EH, Goh RY, Park JI, Han JY. Platelet Activation: The Mechanisms and Potential Biomarkers. Biomed Res Int. 2016;2016:9060143.
255. Rosa JP, George JN, Bainton DF, Nurden AT, Caen JP, McEver RP. Gray platelet syndrome. Demonstration of alpha granule membranes that can fuse with the cell surface. J Clin Invest. 1987;80(4):1138-46.
256. Furlan M. Von Willebrand factor: molecular size and functional activity. Ann Hematol. 1996;72(6):341-8.
257. Mosher DF, Doyle MJ, Jaffe EA. Synthesis and secretion of thrombospondin by cultured human endothelial cells. J Cell Biol. 1982;93(2):343-8.
258. Raugi GJ, Mumby SM, Abbott-Brown D, Bornstein P. Thrombospondin: synthesis and secretion by cells in culture. J Cell Biol. 1982;95(1):351-4.
259. Lawler JW, Slayter HS, Coligan JE. Isolation and characterization of a high molecular weight glycoprotein from human blood platelets. J Biol Chem. 1978;253(23):8609-16.
260. Ferrari DM, Söling HD. The protein disulphide-isomerase family: unravelling a string of folds. Biochem J. 1999;339 (Pt 1)(Pt 1):1-10.

261. Xie L, Chesterman CN, Hogg PJ. Control of von Willebrand factor multimer size by thrombospondin-1. J Exp Med. 2001;193(12):1341-9.
262. Wang A, Liu F, Dong N, Ma Z, Zhang J, Su J, et al. Thrombospondin-1 and ADAMTS13 competitively bind to VWF A2 and A3 domains in vitro. Thromb Res. 2010;126(4):e260-5.
263. Pimanda JE, Ganderton T, Maekawa A, Yap CL, Lawler J, Kershaw G, et al. Role of thrombospondin-1 in control of von Willebrand factor multimer size in mice. J Biol Chem. 2004;279(20):21439-48.
264. Prakash P, Kulkarni PP, Chauhan AK. Thrombospondin 1 requires von Willebrand factor to modulate arterial thrombosis in mice. Blood. 2015;125(2):399-406.
265. Kuijpers MJ, de Witt S, Nergiz-Unal R, van Kruchten R, Korporaal SJ, Verhamme P, et al. Supporting roles of platelet thrombospondin-1 and CD36 in thrombus formation on collagen. Arterioscler Thromb Vasc Biol. 2014;34(6):1187-92.
266. Savage B, Sixma JJ, Ruggeri ZM. Functional self-association of von Willebrand factor during platelet adhesion under flow. Proc Natl Acad Sci U S A. 2002;99(1):425-30.
267. Solecka BA, Weise C, Fuchs B, Kannicht C. Free thiol groups in von Willebrand factor (VWF) are required for its full function under physiological flow conditions. Thromb Res. 2016;137:202-10.
268. Choi H, Aboulfatova K, Pownall HJ, Cook R, Dong JF. Shear-induced disulfide bond formation regulates adhesion activity of von Willebrand factor. J Biol Chem. 2007;282(49):35604-11.
269. Shankaran H, Alexandridis P, Neelamegham S. Aspects of hydrodynamic shear regulating shear-induced platelet activation and self-association of von Willebrand factor in suspension. Blood. 2003;101(7):2637-45.
270. Barg A, Ossig R, Goerge T, Schneider MF, Schillers H, Oberleithner H, et al. Soluble plasma-derived von Willebrand factor assembles to a haemostatically active filamentous network. Thromb Haemost. 2007;97(4):514-26.
271. Dayananda KM, Singh I, Mondal N, Neelamegham S. von Willebrand factor self-association on platelet GpIbalpha under hydrodynamic shear: effect on shear-induced platelet activation. Blood. 2010;116(19):3990-8.
272. Li Y, Choi H, Zhou Z, Nolasco L, Pownall HJ, Voorberg J, et al. Covalent regulation of ULVWF string formation and elongation on endothelial cells under flow conditions. J Thromb Haemost. 2008;6(7):1135-43.
273. Schooten CJ, Tjernberg P, Westein E, Terraube V, Castaman G, Mourik JA, et al. Cysteine-mutations in von Willebrand factor associated with increased clearance. J Thromb Haemost. 2005;3(10):2228-37.

274. Ganderton T, Wong JW, Schroeder C, Hogg PJ. Lateral self-association of VWF involves the Cys2431-Cys2453 disulfide/dithiol in the C2 domain. Blood. 2011;118(19):5312-8.
275. Xie L, Chesterman CN, Hogg PJ. Reduction of von Willebrand factor by endothelial cells. Thromb Haemost. 2000;84(3):506-13.
276. Yeh HC, Zhou Z, Choi H, Tekeoglu S, May W, 3rd, Wang C, et al. Disulfide bond reduction of von Willebrand factor by ADAMTS-13. J Thromb Haemost. 2010;8(12):2778-88.
277. O'Brien HER, Zhang XF, Sanz-Hernandez M, Chion A, Shapiro S, Mobayen G, et al. Blocking von Willebrand factor free thiols inhibits binding to collagen under high and pathological shear stress. J Thromb Haemost. 2020.
278. Sakariassen KS, Orning L, Turitto VT. The impact of blood shear rate on arterial thrombus formation. Future Sci OA. 2015;1(4):Fso30.
279. Zhang C, Kelkar A, Neelamegham S. von Willebrand factor self-association is regulated by the shear-dependent unfolding of the A2 domain. Blood Adv. 2019;3(7):957-68.
280. Furlan M, Robles R, Lämmle B. Partial purification and characterization of a protease from human plasma cleaving von Willebrand factor to fragments produced by in vivo proteolysis. Blood. 1996;87(10):4223-34.
281. Crawley JT, de Groot R, Xiang Y, Luken BM, Lane DA. Unraveling the scissile bond: how ADAMTS13 recognizes and cleaves von Willebrand factor. Blood. 2011;118(12):3212-21.
282. Turner N, Nolasco L, Tao Z, Dong JF, Moake J. Human endothelial cells synthesize and release ADAMTS-13. J Thromb Haemost. 2006;4(6):1396-404.
283. Lian EC. Pathogenesis of thrombotic thrombocytopenic purpura: ADAMTS13 deficiency and beyond. Semin Thromb Hemost. 2005;31(6):625-32.
284. Bongers TN, de Bruijne EL, Dippel DW, de Jong AJ, Deckers JW, Poldermans D, et al. Lower levels of ADAMTS13 are associated with cardiovascular disease in young patients. Atherosclerosis. 2009;207(1):250-4.
285. Sonneveld MA, de Maat MP, Portegies ML, Kavousi M, Hofman A, Turecek PL, et al. Low ADAMTS13 activity is associated with an increased risk of ischemic stroke. Blood. 2015;126(25):2739-46.
286. Feys HB, Vandeputte N, Palla R, Peyvandi F, Peerlinck K, Deckmyn H, et al. Inactivation of ADAMTS13 by plasmin as a potential cause of thrombotic thrombocytopenic purpura. J Thromb Haemost. 2010;8(9):2053-62.
287. Ono T, Mimuro J, Madoiwa S, Soejima K, Kashiwakura Y, Ishiwata A, et al. Severe secondary deficiency of von Willebrand factor-cleaving protease (ADAMTS13) in patients with sepsis-induced disseminated intravascular coagulation: its correlation with development of renal failure. Blood. 2006;107(2):528-34.

288. Zhang X, Halvorsen K, Zhang CZ, Wong WP, Springer TA. Mechanoenzymatic cleavage of the ultralarge vascular protein von Willebrand factor. Science. 2009;324(5932):1330-4.
289. Pipe SW, Montgomery RR, Pratt KP, Lenting PJ, Lillicrap D. Life in the shadow of a dominant partner: the FVIII-VWF association and its clinical implications for hemophilia A. Blood. 2016;128(16):2007-16.
290. Gardner MD, Chion CK, de Groot R, Shah A, Crawley JT, Lane DA. A functional calcium-binding site in the metalloprotease domain of ADAMTS13. Blood. 2009;113(5):1149-57.
291. Feys HB, Anderson PJ, Vanhoorelbeke K, Majerus EM, Sadler JE. Multi-step binding of ADAMTS-13 to von Willebrand factor. J Thromb Haemost. 2009;7(12):2088-95.
292. Banno F, Chauhan AK, Kokame K, Yang J, Miyata S, Wagner DD, et al. The distal carboxyl-terminal domains of ADAMTS13 are required for regulation of in vivo thrombus formation. Blood. 2009;113(21):5323-9.
293. Zhang Q, Zhou YF, Zhang CZ, Zhang X, Lu C, Springer TA. Structural specializations of A2, a force-sensing domain in the ultralarge vascular protein von Willebrand factor. Proc Natl Acad Sci U S A. 2009;106(23):9226-31.
294. Luken BM, Winn LY, Emsley J, Lane DA, Crawley JT. The importance of vicinal cysteines, C1669 and C1670, for von Willebrand factor A2 domain function. Blood. 2010;115(23):4910-3.
295. Kokame K, Matsumoto M, Fujimura Y, Miyata T. VWF73, a region from D1596 to R1668 of von Willebrand factor, provides a minimal substrate for ADAMTS-13. Blood. 2004;103(2):607-12.
296. Gao W, Anderson PJ, Majerus EM, Tuley EA, Sadler JE. Exosite interactions contribute to tension-induced cleavage of von Willebrand factor by the antithrombotic ADAMTS13 metalloprotease. Proc Natl Acad Sci U S A. 2006;103(50):19099-104.
297. Shim K, Anderson PJ, Tuley EA, Wiswall E, Sadler JE. Platelet-VWF complexes are preferred substrates of ADAMTS13 under fluid shear stress. Blood. 2008;111(2):651-7.
298. Davis AK, Makar RS, Stowell CP, Kuter DJ, Dzik WH. ADAMTS13 binds to CD36: a potential mechanism for platelet and endothelial localization of ADAMTS13. Transfusion. 2009;49(2):206-13.
299. Starke RD, Ferraro F, Paschalaki KE, Dryden NH, McKinnon TA, Sutton RE, et al. Endothelial von Willebrand factor regulates angiogenesis. Blood. 2011;117(3):1071-80.
300. Terraube V, Marx I, Denis CV. Role of von Willebrand factor in tumor metastasis. Thromb Res. 2007;120 Suppl 2:S64-70.
301. van Hinsbergh VW. Endothelium--role in regulation of coagulation and inflammation. Semin Immunopathol. 2012;34(1):93-106.

302. Pober JS, Sessa WC. Evolving functions of endothelial cells in inflammation. Nat Rev Immunol. 2007;7(10):803-15.
303. Borchiellini A, Fijnvandraat K, ten Cate JW, Pajkrt D, van Deventer SJ, Pasterkamp G, et al. Quantitative analysis of von Willebrand factor propeptide release in vivo: effect of experimental endotoxemia and administration of 1-deamino-8-D-arginine vasopressin in humans. Blood. 1996;88(8):2951-8.
304. Vischer UM, Ingerslev J, Wollheim CB, Mestries JC, Tsakiris DA, Haefeli WE, et al. Acute von Willebrand factor secretion from the endothelium in vivo: assessment through plasma propeptide (vWf:AgII) Levels. Thromb Haemost. 1997;77(2):387-93.
305. van Mourik JA, Boertjes R, Huisveld IA, Fijnvandraat K, Pajkrt D, van Genderen PJ, et al. von Willebrand factor propeptide in vascular disorders: A tool to distinguish between acute and chronic endothelial cell perturbation. Blood. 1999;94(1):179-85.
306. Vestweber D, Blanks JE. Mechanisms that regulate the function of the selectins and their ligands. Physiol Rev. 1999;79(1):181-213.
307. Springer TA. Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm. Cell. 1994;76(2):301-14.
308. Blann AD, Nadar SK, Lip GY. The adhesion molecule P-selectin and cardiovascular disease. Eur Heart J. 2003;24(24):2166-79.
309. Burger PC, Wagner DD. Platelet P-selectin facilitates atherosclerotic lesion development. Blood. 2003;101(7):2661-6.
310. Semenov AV, Romanov YA, Loktionova SA, Tikhomirov OY, Khachikian MV, Vasil'ev SA, et al. Production of soluble P-selectin by platelets and endothelial cells. Biochemistry (Mosc). 1999;64(11):1326-35.
311. Blann AD, Lip GY. Hypothesis: is soluble P-selectin a new marker of platelet activation? Atherosclerosis. 1997;128(2):135-8.
312. Blann AD, Lip GY, Beevers DG, McCollum CN. Soluble P-selectin in atherosclerosis: a comparison with endothelial cell and platelet markers. Thromb Haemost. 1997;77(6):1077-80.
313. Mayadas TN, Johnson RC, Rayburn H, Hynes RO, Wagner DD. Leukocyte rolling and extravasation are severely compromised in P selectin-deficient mice. Cell. 1993;74(3):541-54.
314. Methia N, André P, Denis CV, Economopoulos M, Wagner DD. Localized reduction of atherosclerosis in von Willebrand factor-deficient mice. Blood. 2001;98(5):1424-8.
315. Komarova YA, Kruse K, Mehta D, Malik AB. Protein Interactions at Endothelial Junctions and Signaling Mechanisms Regulating Endothelial Permeability. Circ Res. 2017;120(1):179-206.

316. Suidan GL, Brill A, De Meyer SF, Voorhees JR, Cifuni SM, Cabral JE, et al. Endothelial Von Willebrand factor promotes blood-brain barrier flexibility and provides protection from hypoxia and seizures in mice. Arterioscler Thromb Vasc Biol. 2013;33(9):2112-20.
317. Su G, Atakilit A, Li JT, Wu N, Bhattacharya M, Zhu J, et al. Absence of integrin αvβ3 enhances vascular leak in mice by inhibiting endothelial cortical actin formation. Am J Respir Crit Care Med. 2012;185(1):58-66.
318. Hillgruber C, Steingräber AK, Pöppelmann B, Denis CV, Ware J, Vestweber D, et al. Blocking von Willebrand factor for treatment of cutaneous inflammation. J Invest Dermatol. 2014;134(1):77-86.
319. Petri B, Broermann A, Li H, Khandoga AG, Zarbock A, Krombach F, et al. von Willebrand factor promotes leukocyte extravasation. Blood. 2010;116(22):4712-9.
320. Zhu X, Cao Y, Wei L, Cai P, Xu H, Luo H, et al. von Willebrand factor contributes to poor outcome in a mouse model of intracerebral haemorrhage. Sci Rep. 2016;6:35901.
321. De Meyer GR, Hoylaerts MF, Kockx MM, Yamamoto H, Herman AG, Bult H. Intimal deposition of functional von Willebrand factor in atherogenesis. Arterioscler Thromb Vasc Biol. 1999;19(10):2524-34.
322. Koivunen E, Ranta TM, Annila A, Taube S, Uppala A, Jokinen M, et al. Inhibition of beta(2) integrin-mediated leukocyte cell adhesion by leucine-leucine-glycine motif-containing peptides. J Cell Biol. 2001;153(5):905-16.
323. Pendu R, Terraube V, Christophe OD, Gahmberg CG, de Groot PG, Lenting PJ, et al. P-selectin glycoprotein ligand 1 and beta2-integrins cooperate in the adhesion of leukocytes to von Willebrand factor. Blood. 2006;108(12):3746-52.
324. McEver RP. Adhesive interactions of leukocytes, platelets, and the vessel wall during hemostasis and inflammation. Thromb Haemost. 2001;86(3):746-56.
325. Adam F, Casari C, Prévost N, Kauskot A, Loubière C, Legendre P, et al. A genetically-engineered von Willebrand disease type 2B mouse model displays defects in hemostasis and inflammation. Sci Rep. 2016;6:26306.
326. Semple JW, Italiano JE, Jr., Freedman J. Platelets and the immune continuum. Nat Rev Immunol. 2011;11(4):264-74.
327. Dhanesha N, Prakash P, Doddapattar P, Khanna I, Pollpeter MJ, Nayak MK, et al. Endothelial Cell-Derived von Willebrand Factor Is the Major Determinant That Mediates von Willebrand Factor-Dependent Acute Ischemic Stroke by Promoting Postischemic Thrombo-Inflammation. Arterioscler Thromb Vasc Biol. 2016;36(9):1829-37.
328. Browder T, Folkman J, Pirie-Shepherd S. The hemostatic system as a regulator of angiogenesis. J Biol Chem. 2000;275(3):1521-4.

329. Inbal A, Dardik R. Role of coagulation factor XIII (FXIII) in angiogenesis and tissue repair. Pathophysiol Haemost Thromb. 2006;35(1-2):162-5.
330. Mackman N. Role of tissue factor in hemostasis, thrombosis, and vascular development. Arterioscler Thromb Vasc Biol. 2004;24(6):1015-22.
331. Tímár J, Döme B, Fazekas K, Janovics A, Paku S. Angiogenesis-dependent diseases and angiogenesis therapy. Pathol Oncol Res. 2001;7(2):85-94.
332. Randi AM, Laffan MA. Von Willebrand factor and angiogenesis: basic and applied issues. J Thromb Haemost. 2017;15(1):13-20.
333. Randi AM, Laffan MA, Starke RD. Von Willebrand factor, angiodysplasia and angiogenesis. Mediterr J Hematol Infect Dis. 2013;5(1):e2013060.
334. Hodivala-Dilke K. alphavbeta3 integrin and angiogenesis: a moody integrin in a changing environment. Curr Opin Cell Biol. 2008;20(5):514-9.
335. Denis C, Williams JA, Lu X, Meyer D, Baruch D. Solid-phase von Willebrand factor contains a conformationally active RGD motif that mediates endothelial cell adhesion through the alpha v beta 3 receptor. Blood. 1993;82(12):3622-30.
336. Mahabeleshwar GH, Feng W, Phillips DR, Byzova TV. Integrin signaling is critical for pathological angiogenesis. J Exp Med. 2006;203(11):2495-507.
337. Soldi R, Mitola S, Strasly M, Defilippi P, Tarone G, Bussolino F. Role of alphavbeta3 integrin in the activation of vascular endothelial growth factor receptor-2. Embo j. 1999;18(4):882-92.
338. Xu H, Cao Y, Yang X, Cai P, Kang L, Zhu X, et al. ADAMTS13 controls vascular remodeling by modifying VWF reactivity during stroke recovery. Blood. 2017;130(1):11-22.
339. Scheppke L, Murphy EA, Zarpellon A, Hofmann JJ, Merkulova A, Shields DJ, et al. Notch promotes vascular maturation by inducing integrin-mediated smooth muscle cell adhesion to the endothelial basement membrane. Blood. 2012;119(9):2149-58.
340. McKinnon TAJ, Starke RD, Ediriwickrema K, Randi AM, Laffan M. Von Willebrand Factor Binds to the Endothelial Growth Factor Angiopoietin-2 Both within Endothelial Cells and Upon Release From Weibel Palade Bodies. Blood. 2011;118(21):698-.
341. Patmore S, Dhami SPS, O'Sullivan JM. Von Willebrand factor and cancer; metastasis and coagulopathies. J Thromb Haemost. 2020.
342. Mussbacher M, Salzmann M, Brostjan C, Hoesel B, Schoergenhofer C, Datler H, et al. Cell Type-Specific Roles of NF-κB Linking Inflammation and Thrombosis. Front Immunol. 2019;10:85.

343. Palacios-Acedo AL, Mège D, Crescence L, Dignat-George F, Dubois C, Panicot-Dubois L. Platelets, Thrombo-Inflammation, and Cancer: Collaborating With the Enemy. Front Immunol. 2019;10:1805.

344. Suter CM, Hogg PJ, Price JT, Chong BH, Ward RL. Identification and characterisation of a platelet GPIb/V/IX-like complex on human breast cancers: implications for the metastatic process. Jpn J Cancer Res. 2001;92(10):1082-92.

345. Gay LJ, Felding-Habermann B. Contribution of platelets to tumour metastasis. Nat Rev Cancer. 2011;11(2):123-34.

346. Bauer AT, Suckau J, Frank K, Desch A, Goertz L, Wagner AH, et al. von Willebrand factor fibers promote cancer-associated platelet aggregation in malignant melanoma of mice and humans. Blood. 2015;125(20):3153-63.

347. Wang WS, Lin JK, Lin TC, Chiou TJ, Liu JH, Yen CC, et al. Plasma von Willebrand factor level as a prognostic indicator of patients with metastatic colorectal carcinoma. World J Gastroenterol. 2005;11(14):2166-70.

348. Guo R, Yang J, Liu X, Wu J, Chen Y. Increased von Willebrand factor over decreased ADAMTS-13 activity is associated with poor prognosis in patients with advanced non-small-cell lung cancer. J Clin Lab Anal. 2018;32(1).

349. Goerge T, Barg A, Schnaeker EM, Poppelmann B, Shpacovitch V, Rattenholl A, et al. Tumor-derived matrix metalloproteinase-1 targets endothelial proteinase-activated receptor 1 promoting endothelial cell activation. Cancer Res. 2006;66(15):7766-74.

350. Xu Y, Pan S, Liu J, Dong F, Cheng Z, Zhang J, et al. GATA3-induced vWF upregulation in the lung adenocarcinoma vasculature. Oncotarget. 2017;8(66):110517-29.

351. Ishihara J, Ishihara A, Sasaki K, Lee SS, Williford JM, Yasui M, et al. Targeted antibody and cytokine cancer immunotherapies through collagen affinity. Sci Transl Med. 2019;11(487).

352. Liu G, Ren YM. [Effect of von Willebrand factor on the biological characteristics of colorectal cancer cells]. Zhonghua Wei Chang Wai Ke Za Zhi. 2010;13(8):616-9.

353. Yang AJ, Wang M, Wang Y, Cai W, Li Q, Zhao TT, et al. Cancer cell-derived von Willebrand factor enhanced metastasis of gastric adenocarcinoma. Oncogenesis. 2018;7(1):12.

354. Mojiri A, Stoletov K, Carrillo MA, Willetts L, Jain S, Godbout R, et al. Functional assessment of von Willebrand factor expression by cancer cells of non-endothelial origin. Oncotarget. 2017;8(8):13015-29.

355. Bharati KP, Prashanth UR. Von Willebrand disease: an overview. Indian J Pharm Sci. 2011;73(1):7-16.

356. Castaman G, Goodeve A, Eikenboom J. Principles of care for the diagnosis and treatment of von Willebrand disease. Haematologica. 2013;98(5):667-74.

357. Mazurier C, Goudemand J, Hilbert L, Caron C, Fressinaud E, Meyer D. Type 2N von Willebrand disease: clinical manifestations, pathophysiology, laboratory diagnosis and molecular biology. Best Pract Res Clin Haematol. 2001;14(2):337-47.
358. Hampshire DJ, Goodeve AC. The international society on thrombosis and haematosis von Willebrand disease database: an update. Semin Thromb Hemost. 2011;37(5):470-9.
359. Ott HW, Griesmacher A, Schnapka-Koepf M, Golderer G, Sieberer A, Spannagl M, et al. Analysis of von Willebrand factor multimers by simultaneous high- and low-resolution vertical SDS-agarose gel electrophoresis and Cy5-labeled antibody high-sensitivity fluorescence detection. Am J Clin Pathol. 2010;133(2):322-30.
360. Nichols WC, Ginsburg D. von Willebrand disease. Medicine (Baltimore). 1997;76(1):1-20.
361. Castaman G, Bertoncello K, Bernardi M, Eikenboom JC, Budde U, Rodeghiero F. Autosomal recessive von Willebrand disease associated with compound heterozygosity for a novel nonsense mutation (2908 del C) and the missense mutation C2362F: definite evidence for the non-penetrance of the C2362F mutation. Am J Hematol. 2007;82(5):376-80.
362. Berber E. The molecular genetics of von Willebrand disease. Turk J Haematol. 2012;29(4):313-24.
363. Hommais A, Stépanian A, Fressinaud E, Mazurier C, Meyer D, Girma JP, et al. Mutations C1157F and C1234W of von Willebrand factor cause intracellular retention with defective multimerization and secretion. J Thromb Haemost. 2006;4(1):148-57.
364. Hommais A, Stépanian A, Fressinaud E, Mazurier C, Pouymayou K, Meyer D, et al. Impaired dimerization of von Willebrand factor subunit due to mutation A2801D in the CK domain results in a recessive type 2A subtype IID von Willebrand disease. Thromb Haemost. 2006;95(5):776-81.
365. Gaucher C, Diéval J, Mazurier C. Characterization of von Willebrand factor gene defects in two unrelated patients with type IIC von Willebrand disease. Blood. 1994;84(4):1024-30.
366. Schneppenheim R, Brassard J, Krey S, Budde U, Kunicki TJ, Holmberg L, et al. Defective dimerization of von Willebrand factor subunits due to a Cys-> Arg mutation in type IID von Willebrand disease. Proc Natl Acad Sci U S A. 1996;93(8):3581-6.
367. Jacobi PM, Gill JC, Flood VH, Jakab DA, Friedman KD, Haberichter SL. Intersection of mechanisms of type 2A VWD through defects in VWF multimerization, secretion, ADAMTS-13 susceptibility, and regulated storage. Blood. 2012;119(19):4543-53.
368. Lyons SE, Bruck ME, Bowie EJ, Ginsburg D. Impaired intracellular transport produced by a subset of type IIA von Willebrand disease mutations. J Biol Chem. 1992;267(7):4424-30.

369. Englender T, Lattuada A, Mannucci PM, Sadler JE, Inbal A. Analysis of Arg834Gln and Val902Glu type 2A von Willebrand disease mutations: studies with recombinant von Willebrand factor and correlation with patient characteristics. Blood. 1996;87(7):2788-94.
370. Rayes J, Hollestelle MJ, Legendre P, Marx I, de Groot PG, Christophe OD, et al. Mutation and ADAMTS13-dependent modulation of disease severity in a mouse model for von Willebrand disease type 2B. Blood. 2010;115(23):4870-7.
371. Ruggeri ZM, Pareti FI, Mannucci PM, Ciavarella N, Zimmerman TS. Heightened interaction between platelets and factor VIII/von Willebrand factor in a new subtype of von Willebrand's disease. N Engl J Med. 1980;302(19):1047-51.
372. Federici AB, Mannucci PM, Castaman G, Baronciani L, Bucciarelli P, Canciani MT, et al. Clinical and molecular predictors of thrombocytopenia and risk of bleeding in patients with von Willebrand disease type 2B: a cohort study of 67 patients. Blood. 2009;113(3):526-34.
373. Ozeki M, Kunishima S, Kasahara K, Funato M, Teramoto T, Kaneko H, et al. A family having type 2B von Willebrand disease with an R1306W mutation: Severe thrombocytopenia leads to the normalization of high molecular weight multimers. Thromb Res. 2010;125(2):e17-22.
374. Favaloro EJ, Bonar RA, Mohammed S, Arbelaez A, Niemann J, Freney R, et al. Type 2M von Willebrand disease - more often misidentified than correctly identified. Haemophilia. 2016;22(3):e145-55.
375. Roberts JC, Flood VH. Laboratory diagnosis of von Willebrand disease. Int J Lab Hematol. 2015;37 Suppl 1(Suppl 1):11-7.
376. Nishino M, Girma JP, Rothschild C, Fressinaud E, Meyer D. New variant of von Willebrand disease with defective binding to factor VIII. Blood. 1989;74(5):1591-9.
377. Casonato A, Galletta E, Sarolo L, Daidone V. Type 2N von Willebrand disease: Characterization and diagnostic difficulties. Haemophilia. 2018;24(1):134-40.
378. Sutherland MS, Keeney S, Bolton-Maggs PH, Hay CR, Will A, Cumming AM. The mutation spectrum associated with type 3 von Willebrand disease in a cohort of patients from the north west of England. Haemophilia. 2009;15(5):1048-57.
379. Sadler JE. Low von Willebrand factor: sometimes a risk factor and sometimes a disease. Hematology Am Soc Hematol Educ Program. 2009:106-12.
380. Castaman G, Linari S. Diagnosis and Treatment of von Willebrand Disease and Rare Bleeding Disorders. J Clin Med. 2017;6(4).
381. O'Sullivan JM, Ward S, Lavin M, O'Donnell JS. von Willebrand factor clearance - biological mechanisms and clinical significance. Br J Haematol. 2018;183(2):185-95.

382. James PD, Notley C, Hegadorn C, Leggo J, Tuttle A, Tinlin S, et al. The mutational spectrum of type 1 von Willebrand disease: Results from a Canadian cohort study. Blood. 2007;109(1):145-54.
383. Flood VH, Gill JC, Friedman KD, Bellissimo DB, Haberichter SL, Montgomery RR. Von Willebrand disease in the United States: a perspective from Wisconsin. Semin Thromb Hemost. 2011;37(5):528-34.
384. Sanders YV, Groeneveld D, Meijer K, Fijnvandraat K, Cnossen MH, van der Bom JG, et al. von Willebrand factor propeptide and the phenotypic classification of von Willebrand disease. Blood. 2015;125(19):3006-13.
385. Wohner N, Muczynski V, Mohamadi A, Legendre P, Proulle V, Aymé G, et al. Macrophage scavenger receptor SR-AI contributes to the clearance of von Willebrand factor. Haematologica. 2018;103(4):728-37.
386. Goodeve A, Eikenboom J, Castaman G, Rodeghiero F, Federici AB, Batlle J, et al. Phenotype and genotype of a cohort of families historically diagnosed with type 1 von Willebrand disease in the European study, Molecular and Clinical Markers for the Diagnosis and Management of Type 1 von Willebrand Disease (MCMDM-1VWD). Blood. 2007;109(1):112-21.
387. Corrales I, Catarino S, Ayats J, Arteta D, Altisent C, Parra R, et al. High-throughput molecular diagnosis of von Willebrand disease by next generation sequencing methods. Haematologica. 2012;97(7):1003-7.
388. Driscoll PC, Gronenborn AM, Clore GM. The influence of stereospecific assignments on the determination of three-dimensional structures of proteins by nuclear magnetic resonance spectroscopy. Application to the sea anemone protein BDS-I. FEBS Lett. 1989;243(2):223-33.
389. Neri D, Szyperski T, Otting G, Senn H, Wüthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling. Biochemistry. 1989;28(19):7510-6.
390. Senn H, Werner B, Messerle BA, Weber C, Traber R, Wüthrich K. Stereospecific assignment of the methyl 1H NMR lines of valine and leucine in polypeptides by nonrandom 13C labelling. FEBS Letters. 1989;249(1):113-8.
391. Sedaghat Doost A, Akbari M, Stevens CV, Setiowati AD, Van der Meeren P. A review on nuclear overhauser enhancement (NOE) and rotating-frame overhauser effect (ROE) NMR techniques in food science: Basic principles and applications. Trends in Food Science & Technology. 2019;86:16-24.

392. Shapiro SE, Nowak AA, Wooding C, Birdsey G, Laffan MA, McKinnon TA. The von Willebrand factor predicted unpaired cysteines are essential for secretion. J Thromb Haemost. 2014;12(2):246-54.
393. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. 2014;5:172.
394. Derman AI, Prinz WA, Belin D, Beckwith J. Mutations that allow disulfide bond formation in the cytoplasm of Escherichia coli. Science. 1993;262(5140):1744-7.
395. Stewart EJ, Aslund F, Beckwith J. Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins. Embo j. 1998;17(19):5543-50.
396. Ritz D, Lim J, Reynolds CM, Poole LB, Beckwith J. Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. Science. 2001;294(5540):158-60.
397. Choi H, Kim S, Mukhopadhyay P, Cho S, Woo J, Storz G, et al. Structural basis of the redox switch in the OxyR transcription factor. Cell. 2001;105(1):103-13.
398. Ortenberg R, Gon S, Porat A, Beckwith J. Interactions of glutaredoxins, ribonucleotide reductase, and components of the DNA replication system of Escherichia coli. Proc Natl Acad Sci U S A. 2004;101(19):7439-44.
399. Bessette PH, Aslund F, Beckwith J, Georgiou G. Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. Proc Natl Acad Sci U S A. 1999;96(24):13703-8.
400. Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Fact. 2012;11:56.
401. Prinz WA, Aslund F, Holmgren A, Beckwith J. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm. J Biol Chem. 1997;272(25):15661-7.
402. Berkmen M. Production of disulfide-bonded proteins in Escherichia coli. Protein Expr Purif. 2012;82(1):240-51.
403. Jonda S, Huber-Wunderlich M, Glockshuber R, Mössner E. Complementation of DsbA deficiency with secreted thioredoxin variants reveals the crucial role of an efficient dithiol oxidant for catalyzed protein folding in the bacterial periplasm. Embo j. 1999;18(12):3271-81.
404. Klint JK, Senff S, Saez NJ, Seshadri R, Lau HY, Bende NS, et al. Production of recombinant disulfide-rich venom peptides for structural and functional analysis via expression in the periplasm of E. coli. PLoS One. 2013;8(5):e63865.

405. Guan Y, Zhu Q, Huang D, Zhao S, Jan Lo L, Peng J. An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide. Scientific Reports. 2015;5(1):13370.
406. Alves VS, Pimenta DC, Sattlegger E, Castilho BA. Biophysical characterization of Gir2, a highly acidic protein of Saccharomyces cerevisiae with anomalous electrophoretic behavior. Biochem Biophys Res Commun. 2004;314(1):229-34.
407. García-Ortega L, De los Ríos V, Martínez-Ruiz A, Oñaderra M, Lacadena J, Martínez del Pozo A, et al. Anomalous electrophoretic behavior of a very acidic protein: ribonuclease U2. Electrophoresis. 2005;26(18):3407-13.
408. Grey MJ, Wang C, Palmer AG, 3rd. Disulfide bond isomerization in basic pancreatic trypsin inhibitor: multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling. J Am Chem Soc. 2003;125(47):14324-35.
409. Sebastião RCO, Pacheco CN, Braga JP, Piló-Veloso D. Diffusion coefficient distribution from NMR-DOSY experiments using Hopfield neural network. Journal of Magnetic Resonance. 2006;182(1):22-8.
410. Berliner L. Protein NMR: Modern Techniques and Biomedical Applications2015. 1-185 p.
411. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR. 1999;13(3):289-302.
412. Wishart DS, Sykes BD, Richards FM. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry. 1992;31(6):1647-51.
413. Sharma D, Rajarathnam K. 13C NMR chemical shifts can predict disulfide bond formation. J Biomol NMR. 2000;18(2):165-71.
414. Chen Y, Campbell SL, Dokholyan NV. Deciphering protein dynamics from NMR data using explicit structure sampling and selection. Biophys J. 2007;93(7):2300-6.
415. Osborne A, Teng Q, Miles EW, Phillips RS. Detection of open and closed conformations of tryptophan synthase by 15N-heteronuclear single-quantum coherence nuclear magnetic resonance of bound 1-15N-L-tryptophan. J Biol Chem. 2003;278(45):44083-90.
416. Park SH, Son WS, Mukhopadhyay R, Valafar H, Opella SJ. Phage-induced alignment of membrane proteins enables the measurement and structural analysis of residual dipolar couplings with dipolar waves and lambda-maps. J Am Chem Soc. 2009;131(40):14140-1.
417. Chen K, Tjandra N. The use of residual dipolar coupling in studying proteins by NMR. Top Curr Chem. 2012;326:47-67.

418. Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol. 1998;5(12):1065-74.
419. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol. 1997;273(1):283-98.
420. Nabuurs SB, Spronk CA, Vuister GW, Vriend G. Traditional biomolecular structure determination by NMR spectroscopy allows for major errors. PLoS Comput Biol. 2006;2(2):e9.
421. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins. 2007;66(4):778-95.
422. Buchner L, Güntert P. Increased reliability of nuclear magnetic resonance protein structures by consensus structure bundles. Structure. 2015;23(2):425-34.
423. Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. Biomol Concepts. 2010;1(3-4):271-83.
424. Harrison PM, Sternberg MJ. The disulphide beta-cross: from cystine geometry and clustering to classification of small disulphide-rich protein folds. J Mol Biol. 1996;264(3):603-23.
425. Shukla D, Trout BL. Understanding the synergistic effect of arginine and glutamic acid mixtures on protein solubility. J Phys Chem B. 2011;115(41):11831-9.
426. Marino SM, Gladyshev VN. Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. J Mol Biol. 2010;404(5):902-16.
427. Cruz MA, Handin RI, Wise RJ. The interaction of the von Willebrand factor-A1 domain with platelet glycoprotein Ib/IX. The role of glycosylation and disulfide bonding in a monomeric recombinant A1 domain protein. J Biol Chem. 1993;268(28):21238-45.
428. Mansouritorghabeh H. Clinical and laboratory approaches to hemophilia a. Iran J Med Sci. 2015;40(3):194-205.
429. Moffitt JR, Chemla YR, Smith SB, Bustamante C. Recent advances in optical tweezers. Annu Rev Biochem. 2008;77:205-28.
430. Müller JP, Löf A, Mielke S, Obser T, Bruetzel LK, Vanderlinden W, et al. pH-Dependent Interactions in Dimers Govern the Mechanics and Structure of von Willebrand Factor. Biophys J. 2016;111(2):312-22.
431. Weiner MP, Costa GL, Schoettlin W, Cline J, Mathur E, Bauer JC. Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. Gene. 1994;151(1-2):119-23.
432. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6(3):277-93.

433. Skinner SP, Fogh RH, Boucher W, Ragan TJ, Mureddu LG, Vuister GW. CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. J Biomol NMR. 2016;66(2):111-24.
434. Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. Journal of Magnetic Resonance (1969). 1990;89(3):496-514.
435. Grzesiek S, Bax A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. Journal of Magnetic Resonance (1969). 1992;99(1):201-7.
436. Grzesiek S, Anglister J, Bax A. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in 13C/15N-Enriched Proteins by Isotropic Mixing of 13C Magnetization. Journal of Magnetic Resonance, Series B. 1993;101(1):114-9.
437. Stejskal EO. Use of Spin Echoes in a Pulsed Magnetic-Field Gradient to Study Anisotropic, Restricted Diffusion and Flow. The Journal of Chemical Physics. 1965;43(10):3597-603.
438. Stejskal EO, Tanner JE. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. The Journal of Chemical Physics. 1965;42(1):288-92.
439. Montelione GT, Lyons BA, Emerson SD, Tashiro M. An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. Journal of the American Chemical Society. 1992;114(27):10974-5.
440. Bax A, Clore GM, Gronenborn AM. 1H・1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. Journal of Magnetic Resonance (1969). 1990;88(2):425-31.
441. Shen Y, Bax A. Protein structural information derived from NMR chemical shift with the neural network program TALOS-N. Methods Mol Biol. 2015;1260:17-32.
442. Berjanskii MV, Wishart DS. A simple method to predict protein flexibility using secondary chemical shifts. J Am Chem Soc. 2005;127(43):14970-1.
443. Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, et al. Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1.beta. Biochemistry. 1989;28(15):6150-6.

444. Hansen DF, Yang D, Feng H, Zhou Z, Wiesner S, Bai Y, et al. An exchange-free measure of 15N transverse relaxation: an NMR spectroscopy application to the study of a folding intermediate with pervasive chemical exchange. J Am Chem Soc. 2007;129(37):11468-79.
445. Trott O, Palmer AG, 3rd. R1rho relaxation outside of the fast-exchange limit. J Magn Reson. 2002;154(1):157-60.