Solution structure of the factor VIII binding region on von Willebrand factor

Nuha Shiltagh

Institute of Structural and Molecular Biology

UCL

Thesis submitted for the degree of Doctor of Philosophy

June 2013

Declaration

This thesis describes research undertaken at the Institute of Structural and Molecular Biology, University College London, between January 2011 and March 2013.

I, Nuha Shiltagh, confirm that the work presented in this thesis is my own. Where there has been collaborative work or where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The preservation of haemostatic integrity is secured by the activities of von Willebrand factor (VWF). Upon vascular damage, VWF acts as a molecular bridge facilitating the initial adhesion and aggregation of platelets to the site of vessel injury, and consequently thrombus formation ensues. Furthermore, VWF is the faithful carrier of procoagulant factor VIII in plasma, thereby prolonging its half-life, and efficiently localising FVIII to the incipient platelet plug. The arrest of bleeding and maintenance of blood volume constancy is critically dependant on VWF; as is exemplified by von Willebrand disease (VWD) – the most common inherited bleeding disorder in man, resulting from defective or deficient VWF protein. Much of the function of VWF has been revealed, however, detailed insight into the molecular structure that enables VWF to orchestrate haemostatic processes, in particular FVIII stabilisation in plasma, is lacking. The high resolution NMR structure of the major FVIII binding region (D') on VWF, and the dynamics and flexibility of its substructure, are presented in this thesis. The complex disulphide-bonded D' region is composed of a two subdomain architecture – TIL'E'. Domain TIL' lacks extensive secondary structure, is strikingly dynamic on at least two timescales measured by NMR relaxation experiments, and this region is coincident with the clustering of pathological mutations leading to decreased FVIII binding affinity (type 2N VWD). This indicates that the conformational fluctuations and backbone malleability of domain TIL' collocate with biological activity. In contrast, the structured domain E', is rigid and contains the most commonly occurring and clinically mildest type 2N VWD mutation. These findings provide important insights into VWF·FVIII complex formation and represent a first step towards revealing the molecular basis of the bleeding diathesis type 2N von Willebrand disease.

Acknowledgements

I would like to thank my supervisors, Professor Edward Tuddenham, Dr D. Flemming Hansen, Dr Keith Gomez and Professor Kaila Srai, for their continuous support during the course of this PhD.

Much of the work in this thesis would not have been possible without Dr Lisa Cabrita, Dr John Kirkpatrick and Dr Andrew Osborne, and for whom a thank you is not enough for being the guardians of my PhD. Dr Lisa Cabrita, for her guidance, wisdom and friendship; Dr John Kirkpatrick for his kindness, willingness to always help and invaluable teaching; and Dr Andrew Osborne, from whom one could learn everything needed to be a great scientist.

I am thankful for my friends, Dr Nico Werbeck, the great bantarian of our time and Dr Andrea Sauerwein, who are as brilliant a post docs as anyone could hope for, and I am very lucky to have worked with and learnt from them; Micha Ben Achim Kunze, whose exceptional talent, fantastic sense of humour and big heart, brightened every day; and Simon Greenwood, our wingman, for his sharp wit, inspiring knowledge and for being such a top guy, for whom I wish great things. This PhD would not have been the same without them.

To my friends Annika Weise, who is as intelligent as she is beautiful, and Dr Amy Robertson, a great scientist and one of the loveliest people I know, thank you.

I am grateful for my amazing parents, Khalid and Einas Shiltagh, and my parents-in-law, Charles and Christine Brookfield, for their love and support.

To my wonderful father-in-law, happy 60th birthday!

Lastly, I would like to thank my gorgeous husband, James; the best man I know, and to whom this thesis is dedicated.

Contents

List of Figures

1. Introduction

List of Tables

1. Introduction

Antiquity established the mythic and moral symbolism of blood that to this day subtly endures. The Egyptians saw blood as the carrier of the vital human spirit and would bathe in the liquid as a restorative. It is because blood conveyed strength to the Romans that gladiators were said to have drunk the blood of fallen opponents. Doctors from the medieval to the Victorian era assumed blood to have fantastical powers, draining it to remove evil humors, transfusing it to pacify the deranged. Contemporary culture also attaches great value to blood, with the blood of Christ as among the holiest sacraments, blood libel as the most insidious slander, and the blood-drinking vampire as the most odious demon. The narrative of blood and its metamorphosis from a mystical substance into a vital component of human anatomy has been an undercurrent to the history of medicine and civilisation[1].

1.1 Haemostasis and the coagulation cascade

The cessation of blood loss from a damaged vessel to maintain blood volume constancy, haemostasis, is one of the most important physiological mechanisms. Haemostasis is accomplished by a complex network of intricately choreographed processes that principally involve the endothelium, platelets, and the coagulation cascade, all of which support the dynamic equilibrium that maintains laminar blood flow and protects against the threat of fatal haemorrhage through breaches in vascular integrity[2]. The modern concept of coagulation presented in 1964 as the Cascade model[3,4], unravelled the basic mechanism of one of the most remarkable characteristics of blood - its ability to clot. The coagulation cascade, forming the third arm of the haemostatic process, is a highly sophisticated amplifying series of enzymatic conversions culminating in fibrin deposition and formation of the definitive haemostatic plug (Figure 1.1). This model has since been further refined as the cell based model, describing the highly interwoven array of physical, cellular and biochemical processes that are primed to react, in an explosive manner, to vascular injury and stem blood loss[5]. The pathologic counterparts of haemostasis are thrombosis and bleeding.

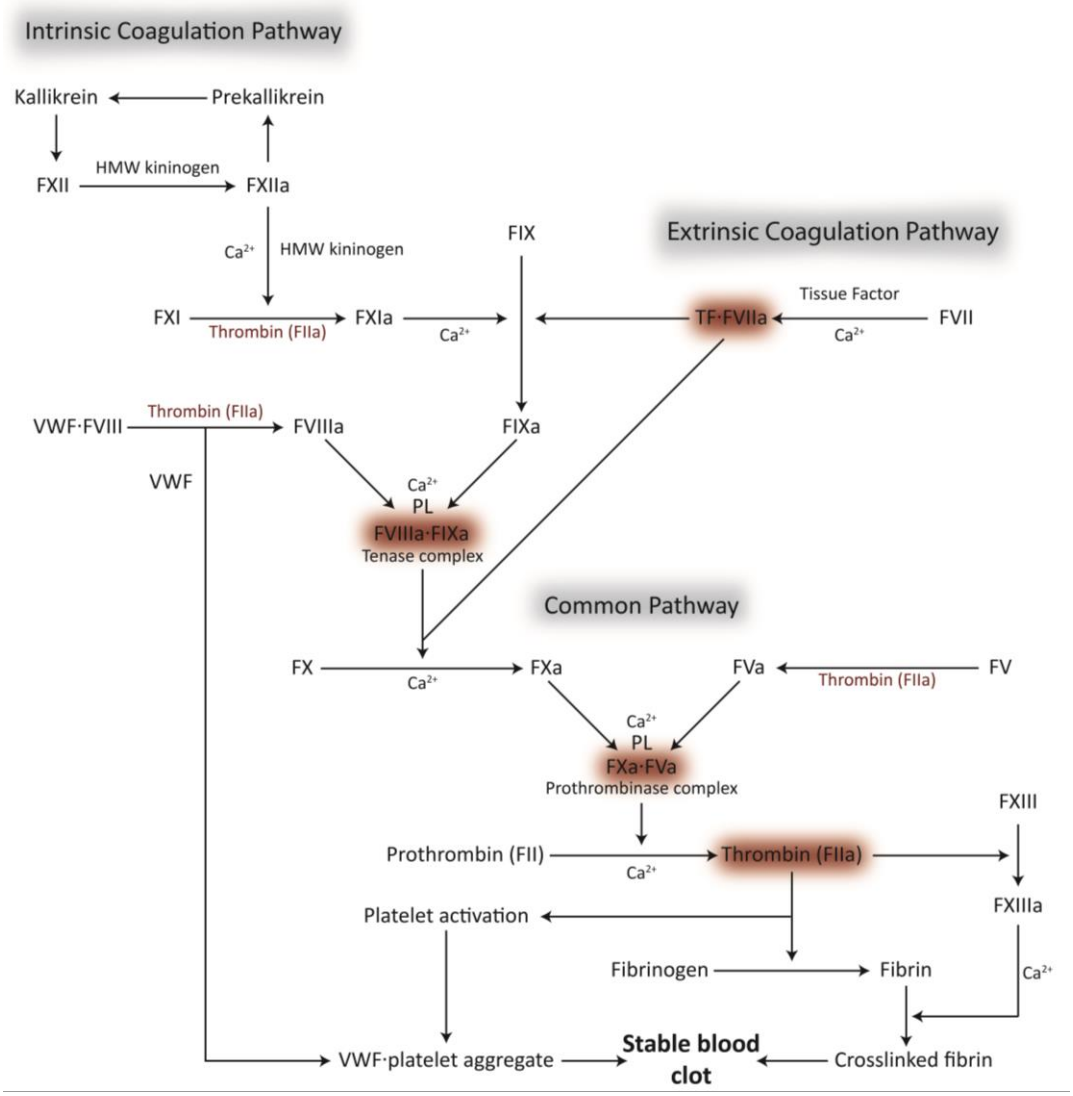


Figure 1.1: Blood coagulation cascade. The coagulation cascade is divided into intrinsic, extrinsic and common pathways. The letter 'a' appended to a coagulation factor indicates that it is activated. Arrows symbolise proteolytic conversion. Cofactors are shown below the arrows. The following abbreviations are used: F, factor; HMW, high molecular weight; TF, tissue factor; PL, phospholipid; Ca^{2+}, calcium. Coagulation is initiated by TF expressing cells released by damaged vascular endothelium and proceeds via the intrinsic pathway. Deficiency of proteins in the extrinsic pathway prolongs a prothrombin time assay. Initiation leads to generation of low levels of thrombin. FVIIa bound to TF activates both FX and FIX. FXa formed by TF·FVIIa binds to FVa and converts small amounts of prothrombin to thrombin. Thrombin generation is then amplified, a major function of which is to activate platelets and convert FVIII into FVIIIa, which acts as a cofactor to FIXa on the platelet surface to support FXa generation. Coagulation is then propagated on the surface of activated platelets via the intrinsic pathway and results in the production of significant levels of thrombin. This leads to fibrin deposition and stabilisation of the platelet clot, forming the definitive haemostatic plug (secondary haemostasis). Deficiency of any intrinsic pathway factors prolongs an activated partial thromboplastin time assay.

The cell based theory divides coagulation into three phases. The initiation phase begins with the extrinsic pathway when contiguity of the vasculature is disturbed and

subendothelial cells (such as smooth muscle cells and fibroblasts) become exposed to flowing blood. These cells expose tissue factor (TF), the principal initiator of coagulation. TF binds and activates coagulation factor VII (FVII), the resultant TF·FVIIa complex proteolytically cleaves FIX and FX into FIXa and FXa, respectively. This allows FXa to associate with cofactor FVa to form the prothrombinase complex on TF expressing cells, which serves to convert prothrombin (FII) into thrombin (Figure 1.1). The slowly accumulating levels of thrombin on TF-bearing cells will further activate platelets that are localised to the site of vessel injury by the activities of von Willebrand factor (VWF). In parallel, thrombin will convert FV to FVa, thus amplifying prothrombinase activity, and convert FVIII to FVIIIa, which acts as a cofactor to FIX on the surface of activated platelets to support FXa generation (Figure 1.1). Additionally, thrombin converts FXI into FXIa. Following the amplification phase, the stage is set for propagation of thrombin generation; this final phase proceeds on the surface of activated platelets mediated by the intrinsic pathway. FIXa, activated during initiation, associates with thrombin-cleaved FVIIIa on the platelet surface (additional FIXa is supplied by platelet-bound FXIa), forming the tenase complex. FVIIIa·FIXa catalyse the conversion of FX to FXa, which rapidly associates with FVa to assemble the prothrombinase complex (Figure 1.1), resulting in a burst of thrombin generation of sufficient magnitude to form an elastic, polymerised fibrin clot[6,7].

1.2 Von Willebrand factor

Preservation of haemostatic integrity is secured by the activities of VWF, synthesised by endothelial cells. VWF is a multimeric plasma sialoglycoprotein that is crucial to primary haemostasis (formation of a platelet clot), particularly in the rapid flow of the arteriolar circulation, a process in which it plays dual roles. Upon vascular damage, VWF acts as a molecular bridge, enabling thrombus formation by facilitating the initial adhesion and aggregation of platelets to the denuded blood vessel wall[8] (Figure 1.2). Furthermore, VWF is the protective carrier of procoagulant FVIII (intrinsic coagulation pathway) in plasma, thereby, prolonging its circulatory half-life and efficiently localising FVIII to the site of the incipient platelet plug, to initiate secondary

haemostasis, concluding in clot stabilisation by fibrin deposition[9] (Figure 1.2). The arrest of bleeding and maintenance of haemodynamic equilibrium is critically dependent upon VWF, as exemplified by von Willebrand disease (VWD); the most common congenital bleeding disorder in man, resulting from defective (type 2) or deficient (type 1 and 3) VWF protein[10,11] and affecting up to 1.3% of the population[12]. VWD is characterised by symptoms of excessive mucocutaneous bleeding such as epistaxis, menorrhagia and prolonged bleeding after trauma. This phenotype is consequent to the two-fold defect in haemostasis, that of impaired platelet adhesion and reduced FVIII concentration[10,13].

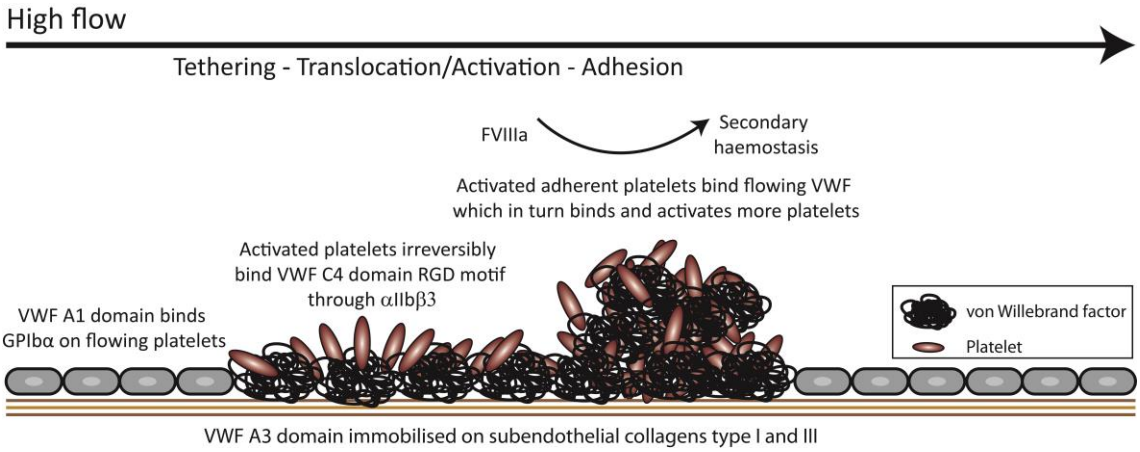


Figure 1.2: Role of von Willebrand factor in thrombus formation. Schematic representation of the mechanisms of platelet aggregation (primary haemostasis) in flowing blood. At the site of vessel injury, VWF is first immobilised on exposed collagen fibrils. Immobilised VWF then binds flowing platelets and following stable platelet-VWF adhesion, platelets are activated and bind more plasma VWF, this forms the substrate onto which additional platelets are recruited and thrombus formation ensues. Upon which co-localised FVIII and tissue factor initiate secondary haemostasis, leading to enmeshment of the platelet plug by polymerised fibrin and formation of an impermeable platelet and fibrin clot. These mechanisms are discussed in detail in section 1.7.

1.3 VWF gene

Cloning of the *VWF* gene was achieved in 1985 using endothelial cell cDNA libraries by four independent groups[14–17]. *VWF* is located distally on the short arm of chromosome 12, at the locus 12p13.3[17,18]. It contains 52 exons, spans 178 kb in length and is transcribed into an 8.7 kb mRNA. Exon 28 is exceptionally large, comprising 1379 base

pairs[19] and is the most mutated region of VWF, encoding domains A1 and A2 that contain essential ligand-binding and cleavage sites, respectively[10] (Figure 1.3).

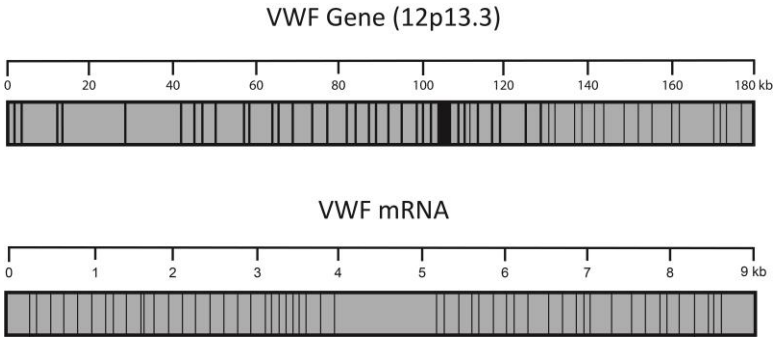


Figure 1.3: Schematic of VWF gene and mRNA. Exons are shown as black vertical bars in the VWF gene. The lower panel shows the VWF mRNA with the large central exon 28, that is the most mutated region of VWF.

Furthermore, studies have also identified an unprocessed 21-29 kb *VWF* pseudogenic sequence (*VWFP*)[20] located on chromosome 22q11.2. *VWFP* partially duplicates exons 23 through 34 of the *VWF* gene, which encode important functional domains in VWF (domains A1-A2-A3). However, the accumulation of missense, nonsense and splice-site mutations in *VWFP* prevents it from generating a functional transcript. The pseudogene, *VWFP*, and *VWF* have diverged only 3.14% in nucleotide sequence (exons, 2.47%; introns, 3.25%), suggesting a recent evolutionary origin – approximately 19-29 million years ago, that is before the separation of orangutan from humans and the African apes (10-15 million years ago), and near the time of divergence of humans and apes from monkeys (20-30 million years ago) by partial gene duplication[21,22]. This is consistent with the presence of *VWFP* in great apes (bonobo, chimpanzee, gorilla and orangutan) but not in more distantly related primates such as the rhesus or spider monkey[23]. The VWF pseudogene may occasionally serve as a reservoir of mutations that can subsequently be introduced into the VWF locus. Potentially pathogenic mutations have been identified in exons 27 and 28 of the VWF gene in VWD patients, and these sequence variations occur consecutively in *VWFP*, therefore it is likely they have been transferred to *VWF* by homologous gene conversion events[24,25].

1.4 VWF expression and tissue distribution

VWF is synthesised throughout the vascular tree, exclusively by endothelial cells[26] and megakaryocytes[27], with the former being the principal source of plasma VWF[28]. Endothelial cell VWF is either constitutively secreted into plasma or stored in cytoplasmic granules, termed Weibel-Palade bodies, that are specific for endothelium[29,30]. Megakaryocyte VWF is stored in the α-granules of its platelet progeny[31,32]. Expression of VWF in the intact endothelium varies both spatially and temporally in different vascular beds. VWF mRNA levels are highest in the lung, spleen aorta and brain endothelial cells, with low levels in the kidney, skeletal muscle, gut and liver[33]. These differences are not accounted for by differences in vascularisation, as the kidney and liver are highly vascularised tissues with relatively low VWF expression, but rather reflect the differential expression of VWF mRNA from tissue to tissue. Thus, in a given tissue, VWF expression is generally greater in the venous side of the circulation than in arterial and arteriolar endothelial cells[34], and greater in the large and small vessels than in the microvasculature[33,34]. This is surprising, as VWF is particularly important in the arterial vessels, which are most directly involved in haemostasis[35]. Furthermore, expression of VWF in the aorta is variable, with the gene product present in clusters of endothelial cells orientated parallel to the longitudinal axis of blood flow and is particularly concentrated in the endothelial cells lining the ostia of the intercostal arteries[36]. Thus, the heterogeneous expression of VWF in endothelial cells from different vessels is an example of vascular differentiation in an organ-specific manner and reflects the transcriptional control of VWF synthesis to maintain the microenvironment-specific haemostatic balance[33,37].

1.5 VWF primary translation product

The primary translation product, preproVWF, is 2813 amino acids (aa) and contains a cleavable 22 aa signal peptide, an unusually large 741 aa propeptide and a mature 2050 aa subunit protein[38,39] (Figure 1.4). A consensus cleavage site for the endoprotease furin separates the propeptide from the remaining domains of mature VWF. The amino acid sequence of VWF is highly conserved. Human VWF is at least 78%

identical to VWF of other placental mammals; 73% identical to opossum VWF; 55-56% identical to bird, reptile and amphibian VWF; and 45-46% identical to fish VWF[40]. Random occurrence of the nucleic acid bases predicts that proteins would contain ~3.3% cysteine[41]. ProVWF however, has an extraordinarily high cysteine content, comprising 324 of the 2813 residues (8.3%), with Cys being the most abundant amino acid. Interestingly, the Cys residues are found to be concentrated in the flanking N- and C-terminal regions, often in clusters of Cys-Cys, Cys-x-Cys or Cys-xx-Cys. The central triplicated A domains are notable for their relative scarcity of cysteines, with only six Cys among 603 residues. It was thought that secreted VWF multimers contain few or no reduced Cys residues, such that successful multimer assembly engages all VWF cysteines in intra- or inter-subunit disulphide bonding[42]. The N-termini at the free ends of the VWF multimer have not engaged in inter-subunit bonding and contain two reduced cysteines[43]. Recent data however, indicates that mature VWF contains unpaired cysteine thiol groups, that are both surface exposed on the protein and buried within the quaternary structure[44–46].

1.5.1 The modular architecture of proVWF

A striking feature of the VWF sequence is its pattern of internal duplication. ProVWF is a highly repetitive protein that is classically thought to comprise four types of homologous domains (A-D), present in two to five copies and encompassing the greater proportion of the protein sequence[15,39,47,48]. A fifth element, 'E', had briefly been proposed[47] based on examination of the mature peptide, however, upon determination of the complete precursor sequence, the 'E' sequences ('E1', aa 788-833; 'E2', aa 2216-2261) were revised as the C-terminal fragments of D domains[49]. The classical D1 and D2 assemblies are tandemly duplicated in the propeptide sequence of the precursor, whereas D3 and D4 are separated by 755 residues: a 3+1 formation. A C-terminal D domain truncation, D' (97 aa), immediately precedes the D3 region (Figure 1.4). Intricate analysis of the proVWF sequence, based on domain family alignments, was used to identify positionally conserved cysteine signatures and to delineate new VWF domain boundaries. This subsequently led to segmentation of the

classical D regions and reassignment of the modular architecture of VWF with respect to five homologous domain repeats: von Willebrand D (VWD), 8-cysteine (C8), trypsin inhibitor-like (TIL), E, and von Willebrand C (VWC) domains[50] (Figure 1.4).

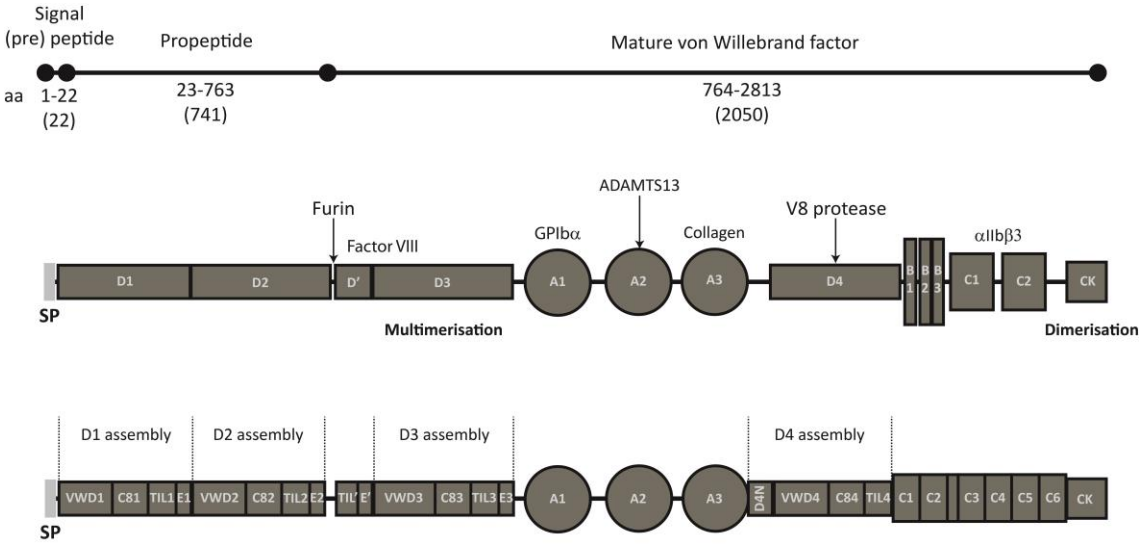


Figure 1.4: The modular architecture of VWF. Upper panel; VWF primary translation product showing the 22 aa signal peptide, a 741 aa propeptide and a 2050 aa mature protein. Middle panel; the classical VWF domain annotation is shown comprising the large VWF D regions (mean length, 359 aa) in a 3+1 formation. Cleavage sites for furin, the metalloprotease ADAMTS13 and staphylococcal protease V8 are indicated with arrows. Binding sites for coagulation factor VIII, platelet glycoprotein Ibα (GPIbα) and platelet integrin receptor αIIbβ3 are indicated. Bottom panel; reassigned VWF domains are shown, whereby each former D region is formed of an assembly or supradomain of VWD-C8-TIL-E domains. A TIL'E' module defines the classical D' region. The B/C regions are revealed to be a tandem arrangement of six VWC domains. An additional domain proximal to the former D4 region has been assigned as D4N[50], and is part of the D4 assembly.

1.5.2 Functional domains of VWF

The functional versatility conferred by domain repeats, occurring in at least 14% of all proteins[51], affords a protein enhanced evolutionary prospects due to the potential for multiple binding and structural roles[52], and it follows that there is a greater abundance of domain repeats in organisms of higher complexity. Many domain families that form repeats are implicated in complex-assembly, cell-adhesion and signalling processes[53]. The VWF subunit contains numerous binding regions, to which specific functions have been ascribed. Binding sites for procoagulant FVIII[54] and, controversially, heparin reside within the N-terminal 272 residues[55,56], however, the indispensable heparin

binding site[57] is located 3' upstream within the A1 domain[58,59], localised to aa 1328-1350[60]. The A1 domain also contains the binding site for platelet GPIbα (residues 1323-1395)[61–63]. Additionally, the A1 domain binds collagen type VI[64,65], which contains VWF type A domains in its noncollagenous regions that may become engaged in homotypic interactions with VWF A1[2]. The principal binding region for fibrillar collagens I and III, however, is located within the A3 domain at aa 1740-1786[66,67]. ADAMTS13 (a disintegrin-like and metalloprotease with thrombospondin type 1 motif, member 13)[68] docks at an exosite within the A2 domain between aa 1660 to 1668[69–71] where it cleaves the VWF Tyr1605-Met1606 peptide bond[72]. Platelet integrin receptor αIIbβ3 binds the VWF RGD tripeptide at aa 2507-2509 in domain VWC4[73,74]. Further VWF binding regions include those for the integral membrane protein P-selectin (aa 776-1241)[75], sulphatides (aa 1392-1423)[76–78] and the snake venom botrocetin (aa 1391 - 1431)[79,80]. Botrocetin from the viper *Bothrops jararaca* enhances the affinity of the A1 domain for platelet receptor GPIbα by increasing the life-time of the A1-GPIbα bond two-fold. This leads to platelet consumption and haemorrhage in the envenomated host[80].

1.5.3 VWF domain evolution

Domain repeats are thought to have evolved through internal gene duplication and recombination events[52] and are most often created by duplication of several domains at a time[53]. In contrast to the evolution of other multidomain proteins that involve the addition of a single domain at either terminus[53,81]. The positions and phase of the intron-exon boundaries within the repeated units of VWF (a detailed analysis of which is given in Appendix 5.1) are consistent with the evolution of these domains by gene segment duplication rather than exon shuffling, for which the 5' and 3' splice junctions of the shuffled exon module are phase compatible with that of its new neighbour[82–84]. In VWF, repeated domains are spread over several exons, the borders of which do not delimit the structural domains and show only limited similarity in the corresponding gene segments that encode the homologous repeats (Figure 1.5). With the exception of the C8-to-TIL linkers that contains coincident phase 1 junctions C-terminal to Cys_{10}

of domains C81 through C84. Overall however, a correlation between the modular architecture and the exon-intron structure is not observed in VWF, although the progression of time may have eroded and obscured the original genomic structure[82], and therefore, the mosaic modularity of VWF alludes to a complex evolutionary history.

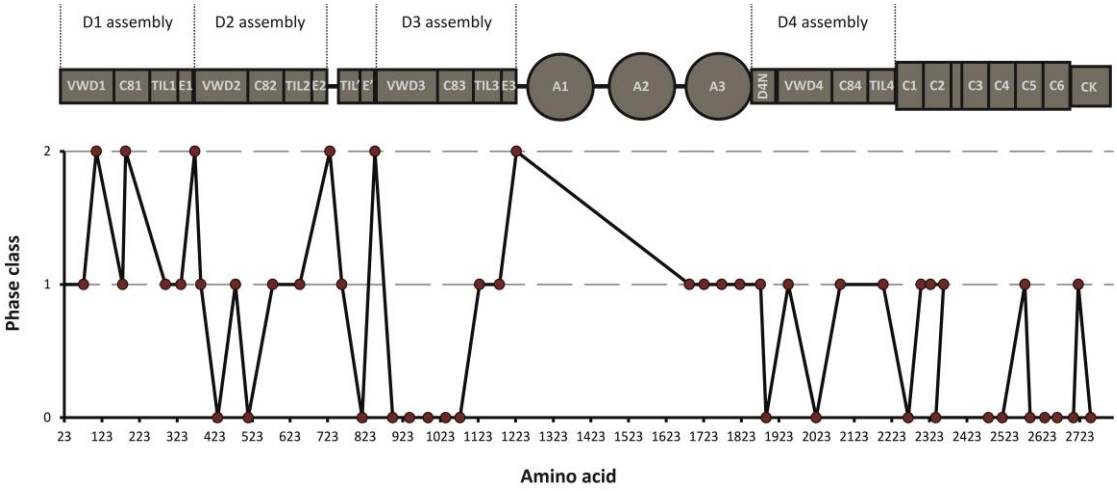


Figure 1.5: Domain boundaries and splice junctions of VWF. ProVWF domain arrangement is shown aligned to the plot. The phase class of the splice junctions, on the y-axis, is plotted against the VWF amino acid sequence. Exon shuffling is governed by splice frame rules. Introns may interrupt the reading frame of a gene between two consecutive codons (phase 0 intro-exon boundary), or between the first and second nucleotide of a codon (phase 1), or between the second and third nucleotide (phase 2). Domains encoded by symmetric exons or symmetric sets of asymmetric exons are likely to have arisen by exon shuffling as their insertion or duplication would not disturb the reading frame[84].

The classical D regions (average length, 359 aa) of VWF constitute a supradomain formed by the assembly of a strongly conserved N- to C-terminal arrangement of VWD-C8-TIL-E domains; a TIL'E' module defines the former D' region. Domains combine in only a limited number of ways, and only a fraction of the possible domain combinations actually exist in genomes[85], implying that this limited repertoire of supradomains that includes VWD-C8-TIL-E, have undergone strong evolutionary selection[81]. A number of proteins contain the VWD-C8-TIL triplet, often in repeats, and are segregated into two groups. VWF, otogelin, mucins and insect hemolectin are evolutionarily related by a triplicate of N-terminally placed VWD-C8-TIL supradomains, followed by a domain characteristic of the respective protein (VWF, triplicated A domains; otogelin, PTS and AbfB domains; mucins, PTS domain; hemolectin, F5/8 type

C domain), and often another VWD-C8-TIL module forming a D1-D2-D3-x-D4 motif (3+1 formation)[86]. They are phylogenetically distinct from other VWD-C8-TIL containing proteins, such as zonadhesin, SCO-spondin, tectorin and IgGFc-binding protein. A common denominator for VWF, hemolectin and the gel-forming mucins is their capacity to form oligomers.

An arrangement of three N-terminally located VWD-C8-TIL modules is involved in the multimerisation of VWF[87], insect hemolectin[88] and mucins[89,90]. The VWF propeptide facilitates multimerisation in the acidic *trans*-Golgi compartments[91] by an oxidoreductase mechanism[92], to which the VWD1 and VWD2 CGLCG motifs are critical[93]. The importance of the triplicate VWD-C8-TIL module in multimerisation indicates that the tendency for the sequential order of the domains to be conserved[94] is likely due to a functional constraint that maintains inter-domain geometry; rather than simply a historical arrangement following a single ancestral recombination event[53,81]. Gel forming mucins appeared early in metazoan evolution, they are found in the starlet sea anemone (*Nematostella vectensis*) and are ancestors to VWF, that did not appear until the vertebrate lineage[86]. This is in alignment with a model of coagulation system evolution[95]. It is interesting that each TIL domain in VWF is followed by either a VWC domain truncation – the E domain repeat - or a VWC domain proper. Analysis of VWD-C8-TIL repeat proteins, reveals that the TIL associated E domain is common and not unique to VWF.

Individual domains of the VWD-C8-TIL-E supradomain, however, are found in more basal species, such that the VWD alone is present in cyanobacteria. The subsequent addition of a C8 domain giving rise to the VWD-C8 doublet occurred in unicellular organisms such as choanoflagellates, the closest group to metazoan organisms, and the assembly of the VWD-C8-TIL triplet, as well as the VWD-C8-TIL-E supradomain appears to have emerged during the multicellularity of animals as a complete VWD-C8-TIL-E sequence was found in the basal placozoan *Trichoplax adhaerens* and cnidarian *Netamostella*, (unpublished observations, Dr Romain Studer).

Reconsideration of the classical VWF B/C domains, following the VWD4-C84-TIL4 module (D4 assembly) indicates that it is formed of six VWC domains, occurring in

tandem and interrupted only by a unique orphan segment between VWC2 and VWC3 (Figure 1.4). This segment may constitute a flexible linker due to the disorder suggested by its compositional bias to threonine, it does however, contain four Cys residues all of which have been chemically shown to form internal disulphide bonds[42], thus hinting at structural order. VWC domains (also referred to as chordin-like cysteine rich repeats) are found in over 500 extracellular proteins, restricted to metazoan genomes. This promiscuous domain is one of the most commonly occurring motifs and is found in diverse protein architectures of variable functions, including the signalling CNN proteins, most isoforms of collagen, thrombospondin, mucins and neuralin[96].

Unlike the lone copy present in the CNN family of proteins, the VWC domain is typically found in multiple copies in other proteins. They are versatile binding units as evidenced by the diversity of processes they functionally support. In most proteins, the VWC repeat acts as an extracellular regulator of bone morphogenic proteins (BMPs) and transforming growth factor-β (TGF-β) signalling. BMPs are potent growth factors, with pleiomorphic functions during development including embryogenesis, skeletal formation, haematopoiesis and neurogenesis[97]. In proteins with multiple VWC domains, each unit can exhibit varying specificities and affinities, there may be silent VWC domains, as well as cooperativity in the function of multiple VWC domains generating higher binding affinities[98,99], suggesting a complicated means of regulation[96]. This behaviour has led to speculation that the disulphide-rich core of VWC provides a scaffold for the attachment of a diversity of loops that can generate different binding specificities similar to an immunoglobulin – making the tandem VWF C domains highly interesting. In the conotoxin superfamily, comprising many hundreds of members with numerous disulphide bonds, a similar principle has been established[100]. The cellular role of most VWC domains however is as yet unknown[101] and the only functional role designated to the VWF C domains is platelet integrin αIIbβ3 binding at VWC4[102].

1.5.4 VWD-C8-TIL-E supradomain

An alignment of the VWF VWD-C8-TIL-E supradomains, highlighting the conserved cysteines and motifs within the domains is shown in Figure 1.6. VWD-C8-TIL triplets in VWF assemblies D1 through D3 are followed by an E domain, whereas assembly D4 is followed by a full VWF C domain (VWC1), the first six cysteines of which align to the E domains. The VWD domain contains four conserved Cys residues; the first N-terminal pair of cysteines are disconnected by 21 amino acids, and flank a central TFDG motif. The second pair of conserved cysteines participate in the formation of the vicinal cysteine residue motif, CGLCG, found in VWD1, VWD2 and VWD3, which corresponds to the CGxC catalytic site of the thioredoxin superfamily (including protein disulphide isomerase)[103]. A truncated motif, GLCG, is present in the corresponding position in VWD4 (Figure 1.6). The C8 and TIL domains are enriched in positionally conserved cysteines; TIL units contain ten conserved Cys residues and participate in the formation of D assemblies, as observed in many extracellular proteins including VWF, or are found alone in serine protease inhibitors, such as potent nematode anticoagulants that are specific inhibitors of FXa and the TF·FVIIa complex[104,105]. The nematode anticoagulant proteins, homologues of the VWF TIL units, are targets for antithrombotic drug development and are currently undergoing clinical and preclinical trials[106].

VWF TIL2 domain contains the tripeptide Arg-Gly-Asp (RGD) – a cell adhesion motif that is also present in domain VWC4 where it forms part of the binding region for platelet integrin receptor αIIbβ3[73,74]. However, the significance of the RGD motif in the prodomains is not known (Figure 1.6). The TIL unit belongs to the *Ascaris* protease inhibitor family (MEROPS clan IA, family I8); designated as a single evolutionary line of inhibitors defined by a conserved protein fold (VWF MEROPS ID: 108.950)[107–109]. The *Ascaris* family predominantly inhibit serine proteases of the S1 family, containing the catalytic triad His-Asp-Ser[110,111], but also M4 family metalloproteases (that does not include ADAMTS13). S1 family inhibition by TIL units is achieved by the Laskowski (standard) mechanism[111,112], whereby an extended solvent-exposed convex combining loop of canonical conformation, surrounds the reactive site P1-P1' scissile bond forming a serine protease substrate analogue[112,113].

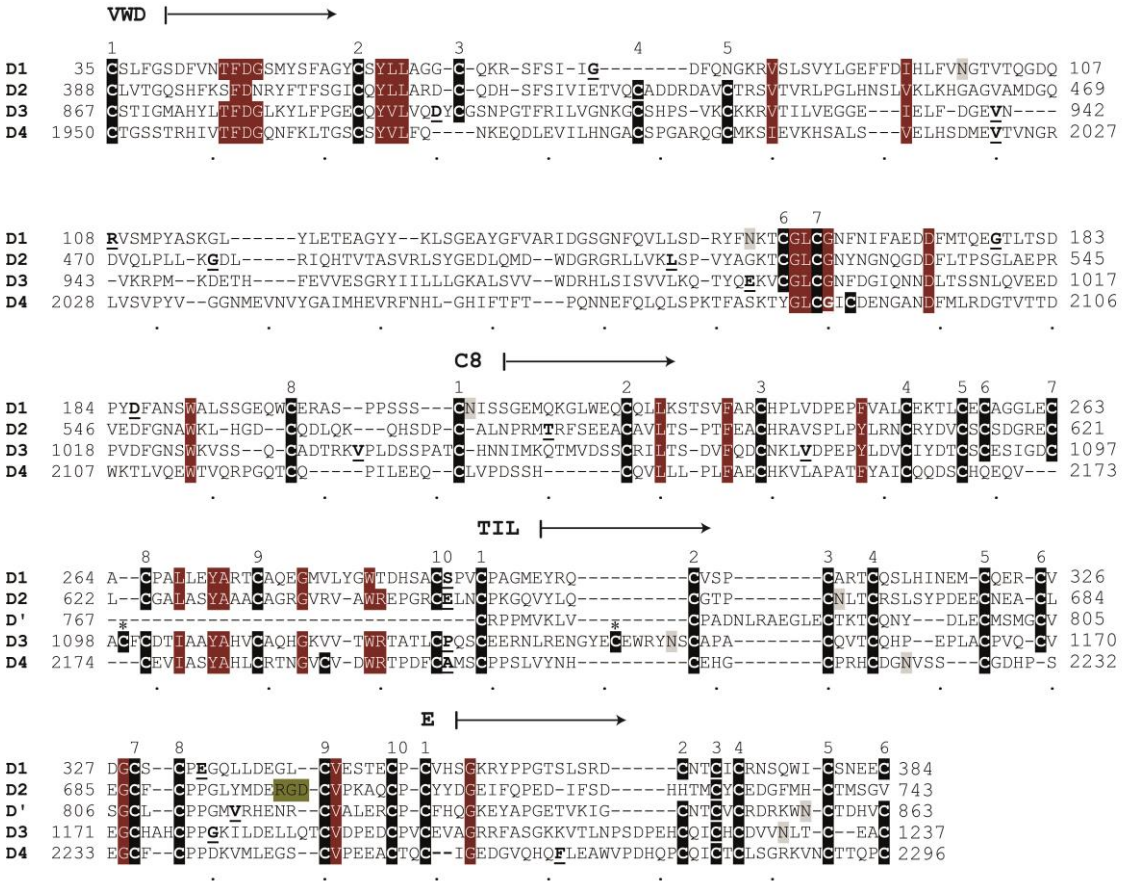


Figure 1.6: The classical D regions are formed of VWD-C8-TIL-E supradomains. Alignment of the VWF supradomains (VWD-C8-TIL-E) formed of the classical D1, D2, D3 and D4 assemblies is shown. Conserved cysteines are highlighted in black and numbered above the alignment according to the order in which they occur in the sequence, such that 1 represents Cys_1. Cys_1-Cys_6 of domain VWF C1 (aa 2257-2296) aligned to the E domains is also shown, to highlight sequence homology. Conserved motifs are highlighted in red; glycosylation sites are highlighted in grey; intron-exon boundaries are shown in bold and underlined. The VWD domain **C**GL**C**G motif is encompassed by Cys_6 and Cys_7. The TIL2 RGD motif is highlighted in green.

The presence of the conserved TIL unit was alluded to early following the cloning of the VWF sequence. Sequences Cys_3 to Cys_{10} of mature VWF TIL' and TIL4 domains, were initially termed 'E1' and 'E2' domains, respectively[47]. This annotation was subsequently made obsolete by enveloping the 'E' domains within the C-terminal D regions[49]. PreproVWF contains five TIL domain units, separated by an average of 425 residues. Interestingly, making it a potential homotypic compound serine protease inhibitor[108]. There are however, no published experimental data as to the possible inhibitory activity of VWF.

The average pairwise sequence identity (and similarity)[114] between the VWD-C8-TIL-E supradomains is 29.4% (42.3%). All pairwise sequence identities are above the twilight

zone limit of 25%[115], thereby allowing their evolutionary relatedness to be inferred. The VWD4-C84-TIL4 triplet was extended to encompass the first six conserved cysteines of VWF domain C1, residues 2257-2296 (Figure 1.6). Pairwise analysis of the sequences between each VWD-C8-TIL-E assembly indicates that VWD4-C84-TIL4-C1 (D4 region) is the most remote module, whereas VWD1-C81-TIL1-E1 (D1) and VWD2-C82-TIL2-E2 (D2) share the highest sequence identity and similarity. Furthermore, domains TIL' and E', forming the classical D' region, are most similar to TIL1 and E1 of region D1, respectively (Figure 1.7).

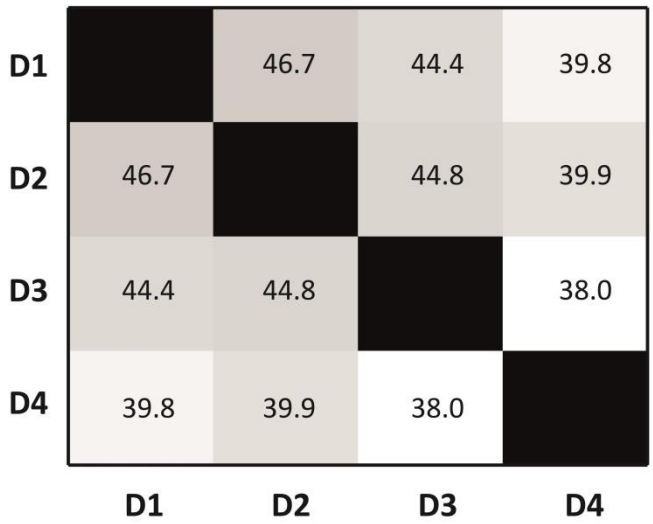


Figure 1.7: Pattern of internal D assembly duplication in VWF. The D4 assembly includes the first six cysteines of domain C1. The intensity of the squares in the matrix reflects the percentage similarity, implementing the NEEDLE algorithm[114], between the D regions (VWD-C8-TIL-E), with darker colour for higher scores. The regions aligned are shown on the x- and y-axes. It is clear that the D4 region is least similar to regions D1 to D3. With thanks to Micha Ben Achim Kunze for generating the matrix.

1.5.5 VWC and E domains

By analysis of homologies to the CNN protein family and thrombospondin, there have been suggestions in the literature that the classical VWF B domains may actually be formed of VWC domains, and that there are a greater number of C domains in VWF than originally reported[116,117]. The VWC repeat is typically ~70 amino acids in length and contains ten conserved Cys residues, forming five disulphide bonds, arranged in a distinctive pattern with the following consensus signature: C-C_2xxC_3xC_4-C-C-C-C_8C_9xC_{10} (connectivity 1-4, 2-8, 3-5, 6-9, 7-10)[101,118]. The motifs C_2xxC_3xC_4 and C_8C_9xC_{10} are highly conserved, as are a glycine and an aromatic residue, often tryptophan, between

the first cysteine pair[116], otherwise the intervening VWC domain sequences contain a high degree of variability (Figure 1.8), wherein the average sequence identity (and similarity) between the VWF C domains is 20.3% (29.0%).

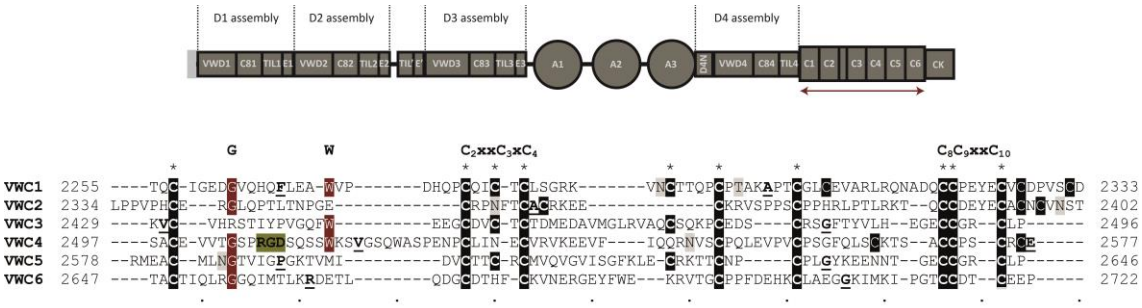


Figure 1.8: VWF C domains. The six tandem VWC domains are aligned below the schematic of the VWF domain arrangement, highlighting the VWC domain region, C1 to C6. In the alignment, conserved cysteines highlighted in black and conserved motifs in red. Glycosylation sites are highlighted in grey and intron-exon borders are shown in bold and underlined. The Arg-Gly-Asp (RGD) sequence in VWC4 that binds platelet integrin receptor αIIbβ3 is highlighted in green. With the exception of positional cysteine conservation, the VWC domain sequence is highly variable.

There are data to suggest that conserved Gly and Pro residues, of which there are clusters throughout the VWF C domains, may act to inhibit aggregation[119–121] and this may be of importance to VWF. The tandem arrangement of six C domains presents a unique folding problem to VWF, as a high local homologous domain concentration makes proteins vulnerable to misfolding and aggregation[122]. In most multidomain proteins, domains are adjacent to non-homologous neighbours, such that misfolding events are unlikely[81]. However, for the VWC region of VWF, which has arisen due to internal duplication[53], misfolding is potentially possible. Adjacent domains in a protein are less similar than non-adjacent domains, whereby, less than 10% of adjacent domain pairs have sequence identities greater than 40%, and the identity between adjacent domains peaks at 30% in eukaryotes[122,123]. Therefore, indicating that adjacent domains have sufficiently diverged in sequence to prevent interdomain misfolding and subsequent aggregation. In VWF, the pressure for divergence is seen in the overall low level of non-cysteine sequence conservation in the C domains; their pairwise identities range from 9.8% (VWC2-VWC3) to a maximum of 29.1% (VWC1-VWC2), which may serve to overcome the high local homologous domain concentration that would otherwise increase the aggregation propensity of VWF[122] (Figure 1.9). Another distinctive evolutionary characteristic of tandem domain repeats is chunk duplication

of multiple domains at a time that prevents identical domains being placed immediately adjacent to one another[81]. Such is the case for the VWF D region supradomains (Figure 1.7) and C domains (Figure 1.9).

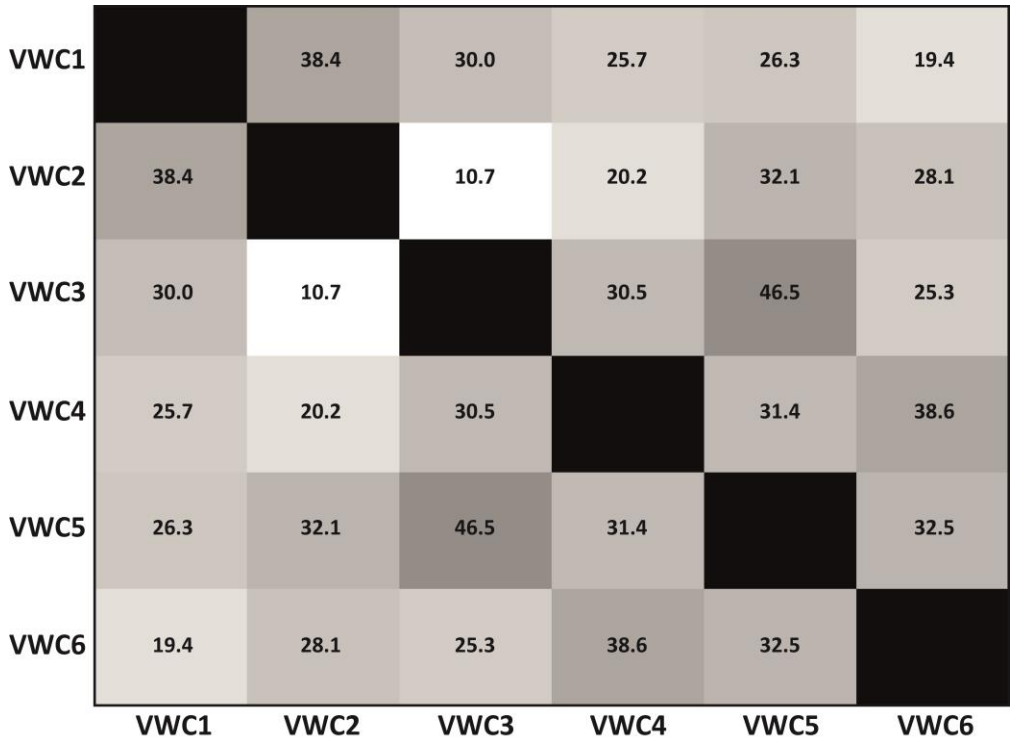


Figure 1.9: Pattern of internal C domain duplication in VWF. The intensity of the squares in the matrix reflects the percentage similarity between the six tandem VWF C domains, with darker colour for higher degrees of similarity. The regions aligned are shown on the x- and y-axes. With thanks to Micha Ben Achim Kunze for generating the matrix.

VWF domain VWC1 has the highest sequence similarity with VWC2 (38.4%), VWC3 is most similar to VWC5 (46.5%) and VWC4 is most similar to VWC6 (38.6%). Therefore, it appears that a unit of two C domains has been duplicated twice in VWF, to give rise to the tandem arrangement of six VWC domains.

The E domains are delineated by their conservation of the first six cysteines of a VWC domain, C_1-C_2xxC_3xC_4-C_5-C_6, and are divergently related to the VWC domains. The presence of a VWC truncation E domain in VWF was suggested soon after determination of the VWF primary sequence. This domain was termed Cx (aa 826-853) and was found to be highly similar to residues 2424-2455 and 2575-2607 of the classical C1 and C2 domains, respectively (Cys_1 to Cys_4 of VWC3 and VWC5)[47,116]. The

recently revised modular architecture of VWF (Figure 1.4) shows that the Cx fragment corresponds to the extreme C-terminus of TIL' and the majority of E' (Figure 1.6). The newly assigned VWF E domains are part of the fibronectin I-like superfamily, formed of disulphide-rich, all-β domains. This domain family includes the FNI module of fibronectin, β-microseminoprotein and subdomain 1 of a VWC module. Strong structural similarity is highly indicative of an evolutionary relationship[124], and this conservation may be due to maintenance of function[101]. The E repeat likely represents an evolutionary isoform of a VWC domain, as a result of alternative splicing[125–127]; consistent with this theory is the intron-exon splice junction present between the sixth and seventh conserved cysteine residues of VWF domain C1 (Figure 1.8 and Appendix 5.1), which immediately succeeds the D4 assembly. It would then simply follow that the E domains evolved by truncation of the random coil subdomain 2 of VWF C1 through alternative splicing, an event without structural ramifications. Whether this is to introduce functional diversity is not clear. That the E domains are repeated within the VWF architecture and are part of the VWD-C8-TIL-E supradomain in the first three D regions, and the remote D4 module precedes a VWC domain true, persuasively suggests that domain VWC1 is the forefather of VWF E domains.

1.5.6 Model of VWF D region evolution

Thus, the ancient gene structure that gave rise to the current VWF D region architecture can be rationalised based on Occam's razor: the remote ancestral gene of the present D assemblies (VWD-C8-TIL-E supradomains) was first duplicated to first give rise to VWD4-C84-TIL4-C1 (D4-C1) and the tandem arrangement of D1-D2; during the evolution of which subdomain 2 of VWF C1 domain was truncated giving rise to VWD1-C81-TIL1-E1 (D1) and VWD2-C82-TIL2-E2 (D2). The next event involved the imperfect duplication of the C-terminus of D1 (TIL1-E1) and the entire D2 region (VWD2-C82-TIL2-E2) to give the D'D3 (TIL'E'-VWD3-C83-TIL3-E3) arrangement. A corresponding tandem arrangement of three D regions is present in basal metazoa[86], and multiple expansions of the VWD-C8-TIL-E supradomain were already present in the ancestor of multicellular animals, therefore the VWF D regions did not duplicate

independently in VWF, but rather are the result of evolutionary descent. This is a unifying model that would account for the sequence composition of the D regions; such that region D4 is the most divergent in sequence, the TIL'E' domains are most similar to the homologous segment in region D1, and that the E domain is a splicing isoform of VWF C1 domain (Figure 1.10).

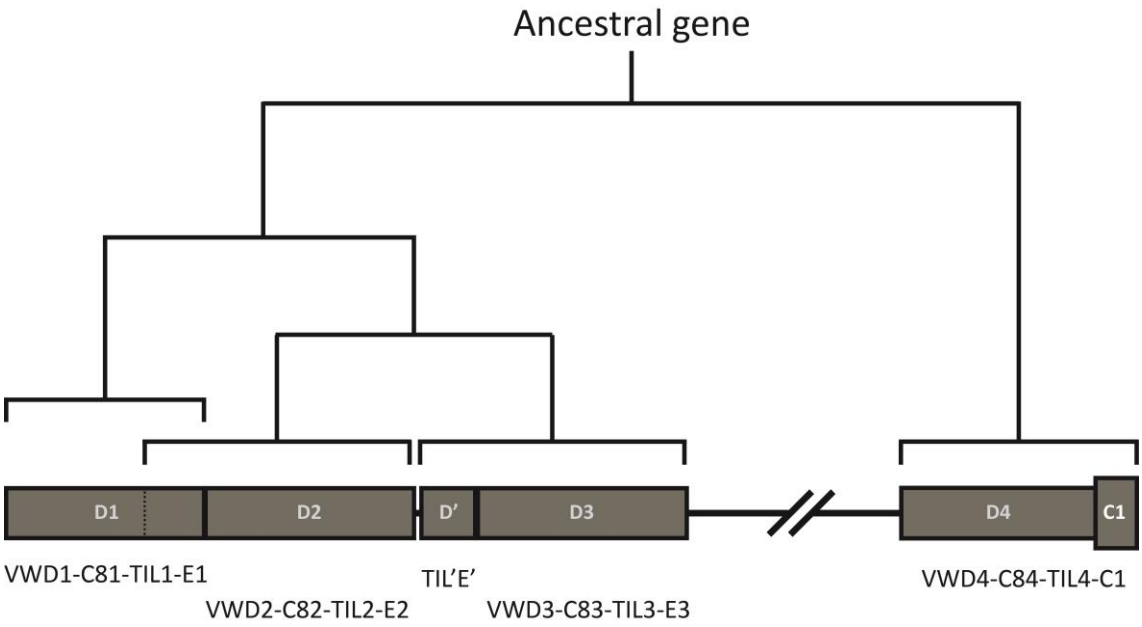


Figure 1.10: Proposed model of D region evolution in VWF. The VWF D regions evolved by gene segment duplication. The ancestral sequence duplicated yielding one sequence that gave rise to region D4 and VWF domain C1, the second sequence evolved to lose subdomain 2 of domain C1 and underwent further duplication. The next such event gave rise to the present tandem arrangement of region D1D2. The most recent duplication involved the C-terminus of region D1 and the entirety of D2, yielding D' (TIL'E') and D3. This model of VWF D region evolution, corresponds to the early model proposed by Shelton *et al*[49] and was further refined and confirmed by phylogenetic analysis under maximum likelihood (Dr Romain Studer, unpublished observations).

1.6 Biosynthesis of VWF

Native VWF is a disulphide-bonded homopolymer, and the multimeric composition of VWF is a critical determinant of its haemostatic potency[128]. Multimer assembly is preceded by the formation of dimeric protomers (~500 kDa)[129] linked near their C-termini, which through N-terminal disulphide-bonding assemble into a logarithmic distribution of multimers[129,130], reaching sizes in excess of 20,000 kDa[131,132]. Polymerisation is facilitated by thiol:disulphide oxidoreductase-like sequences in the propeptide[93]. During synthesis, VWF is extensively glycosylated with N- and O-linked oligosaccharides, whose side chains contribute approximately 20% of the mass of VWF[133] and are believed to affect its functional and structural integrity[134,135].

Early metal-shadowing experiments visualised the gross VWF domain arrangement; VWF constitutively secreted into plasma usually appear as loosely coiled prolate ellipsoids[136], with a modal diameter of 100-150 nm[137], whereas globular VWF undergoes extension under shear flow[138] to form active filamentous polymers that reveal a dimeric periodicity of 120-140 nm[139,140], reaching lengths over 1 μm, and up to the diameter of a platelet[137–139]. The protomer has a two-fold symmetry about a central axis (radius of gyration ~30 nm)[136] and the site of dimerisation is a small C-terminal central node (3.4 × 6.4 nm), connected by long, flexible 34.0 × 2.0 nm fibrillar rods, joined to large globular 26.0 × 6.5 nm ends (N-termini)[136,139]. The large globular regions comprise amino-terminal 2128 residues (up to the D4 assembly) and the rod region is formed of aa 2129-2813 (VWC domains and domain CK). The rod region contains no α-helices and is predominantly composed of β-sheet structure[56]. Cooperativity between the VWD-C8-TIL-E supradomains would support the gross domain arrangement of N-terminal VWF; a large globular formation consisting of smaller rounded structures (Figure 1.11). VWF is the largest known soluble vertebrate protein, and its unusual size poses difficulties in biosynthesis, storage and secretion[140].

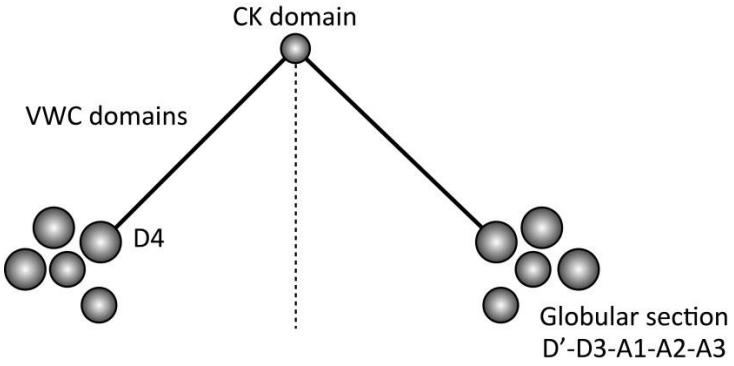


Figure 1.11: Macromolecular structure of VWF. Schematic of the VWF dimeric protomer based on early metal shadowing electron microscopy and small-angle neutron scattering experiments[136,139]. The CK domain is dimeric and is the central axis of two-fold symmetry.

1.6.1 Dimerisation in the endoplasmic reticulum

Following translation of VWF mRNA and translocation into the endoplasmic reticulum (ER), the signal peptide is cleaved. In the ER, proVWF undergoes N-linked glycosylation and dimerisation[141]. Twelve N-linked oligosaccharide units are added to the proVWF molecule at the consensus sequence Asn-x-Ser/Thr (Asn857, Asn1231, Asn1515, Asn1574, Asn2223, Asn2290, Asn2357, Asn2400, Asn2546, Asn2585, Asn2635 and Asn2790)[38], and an additional rare sequon, Asn-x-Cys (Asn1147) is also glycoslylated[142]. Mapping of the VWF N-glycome shows its great complexity as a population of ~300 distinct structures. Complex-type N-glycan chains, predominantly bi-antennary, constitute the dominant structures[135,142,143] (Figure 1.12).

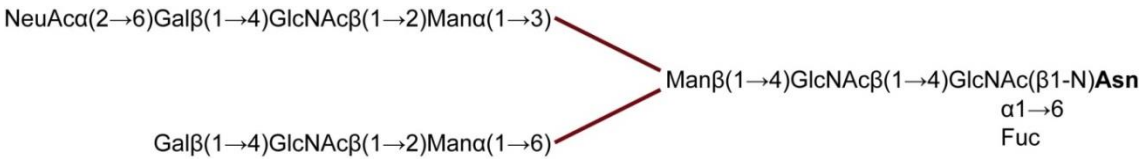


Figure 1.12: Primary structure of the major N-linked glycans on VWF. Man, mannose; GlcNAc, *N*-acetylglucosamine; Gal, galactose; NeuAc, *N*-acetylneuraminic (sialic) acid; Gal(β1→4)GlcNAc, *N*-acetyllactosamine.

These are highly unusual with respect to other plasma glycoproteins as 13% contain blood group A, B and O(H) antigenic structures[143] (α2-macroglobulin also expresses ABO(H) antigens[144]). The greater number of N-glycosylation sites are located in the

vicinity of, or within, a functional site and it is therefore unsurprising that they affect VWF property and function. Blocking the addition of N-linked glycans with tunicamycin abolished VWF dimerisation and subsequent processing and secretion[91], of which Asn99, Asn857 and Asn2790 were of critical importance[145]. Further, it has been shown that glycosylation at site N857 (VWD3 domain) may be essential to formation of the correct disulphide bond topology in region D'D3 of VWF[145] (Figure 1.13).

One of the most striking effects of N-linked glycosylation is the correlation between ABO blood group and plasma VWF levels. There are significant differences in VWF levels between ABO blood groups, with 25% lower VWF levels in blood group O than in non-O individuals (AB>A>B>O)[146,147]. VWF plasma levels are further reduced in persons with the Bombay phenotype, who lack expression of ABO(H) antigens[148]. The precise mechanism by which ABO(H) antigens affect plasma levels remains unclear, however, the overriding evidence suggests an effect on accelerated VWF clearance[142,147]. ABO(H) antigens also alter VWF susceptibility to proteolysis by ADAMTS13, a specific protease regulating VWF multimeric size, with a cleavage rate in the rank order O>B>A>AB[149]. VWF sialylation also specifically enhances its susceptibility to ADAMTS13 cleavage[142,150]. Despite the demonstrated importance of N-linked glycosylation to VWF, only two naturally occurring N-linked glycan site mutations are reported, and neither causes VWD by a direct effect on glycosylation. Mutant Asn1231Thr represents a gene conversion event from the partial psuedogene sequence on chromosome 22[151] and is not thought to be causative for VWD, whereas in mutant Asn2546Tyr there was found to be an absence of mRNA containing this mutation, suggesting that another mutation was present in *cis* which completely prevented expression[145].

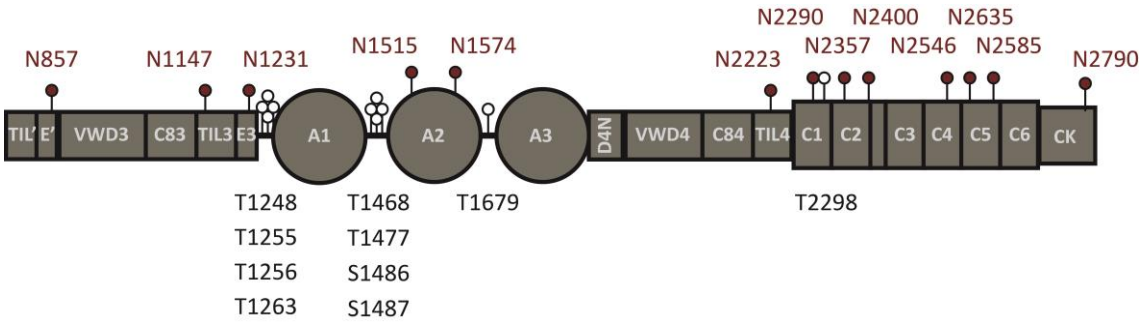


Figure 1.13: The VWF glycome. The mature VWF protein is shown. N- and O-glycosylation sites are represented by closed red and open lollipops, respectively.

ProVWF subunits dimerise through disulphide bonds near their C-termini ('tail to tail' bonding) forming the dimeric protomer. The cysteines involved in dimerisation have been localised to the last 90 residues of VWF that comprise the cysteine knot (CK) domain (aa 2724-2811)[152]. The CK superfamily of proteins, including TGF-β and platelet derived growth factor, are always found at the C-terminus of extracellular proteins, but differ in the position of the cysteine that dimerises two monomers[153]. CK domains include six cysteines arranged in a knot-like topology. Disulphide bonds between the second and fifth cysteine, and between the third and sixth cysteine, form a macrocyclic ring that is penetrated by a disulphide bond between the first and fourth cysteines[152] (Figure 1.14). Chemical assignment resolved eight of the eleven cysteines in VWF CK domain leaving three potential dimerisation candidates: Cys2771, Cys2773 and Cys2811[152]. However, symmetry considerations and electron microscopy imaging support a VWF dimerisation model that is analogous to TGF-β, in which the long axes of the two monomers are anti-parallel, and thus dimerisation through the cysteine that aligns with the TGF-β dimerisation cysteine[50], that is VWF Cys2773. This is further supported by the phenotype of Cys2773Arg/Ser mutants that results in aberrant C-terminal linked dimers, the consequence of which is a qualitative multimerisation defect, VWD type 2A (IID)[154,155]. Exit from the ER is regulated by both glycosylation and dimerisation[23].

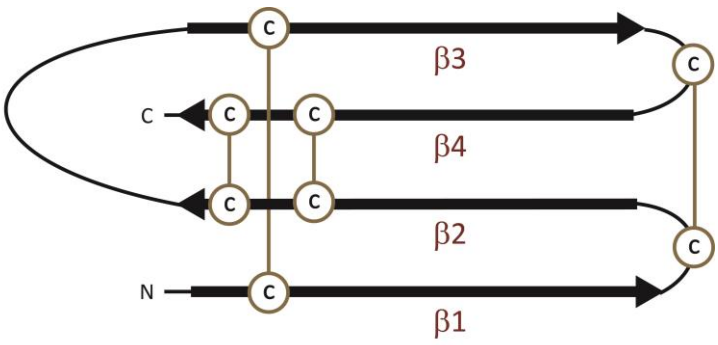


Figure 1.14: CK domain knot-like topology. Schematic of a representative CK domain, showing the knot-like disulphide bond topology. Cysteine residues are designated C and disulphide bonds are shown by gold connecting bars.

1.6.2 Multimer assembly in the Golgi

The intracellular assembly of VWF presents a challenge to the cell, as enormous multimers are formed and these may exceed the capacity for vesicular transport from the ER to the Golgi, whereas, there seems to be no obvious size limit for transport form the late-Golgi to the cell surface[156]. Compatible with these size limitations, VWF has evolved a unique two-stage mechanism of assembly whereby, dimerisation occurs in the ER and the multimerisation event is postponed until the Golgi compartments (Figure 1.15). However, the ER has chaperones and a redox potential suitable for disulphide bond formation. In contrast, the Golgi lacks known chaperones and is relatively acidic, which generally prevents disulphide rearrangement. Therefore, VWF requires a specific mechanism to promote multimer formation in the Golgi apparatus[156].

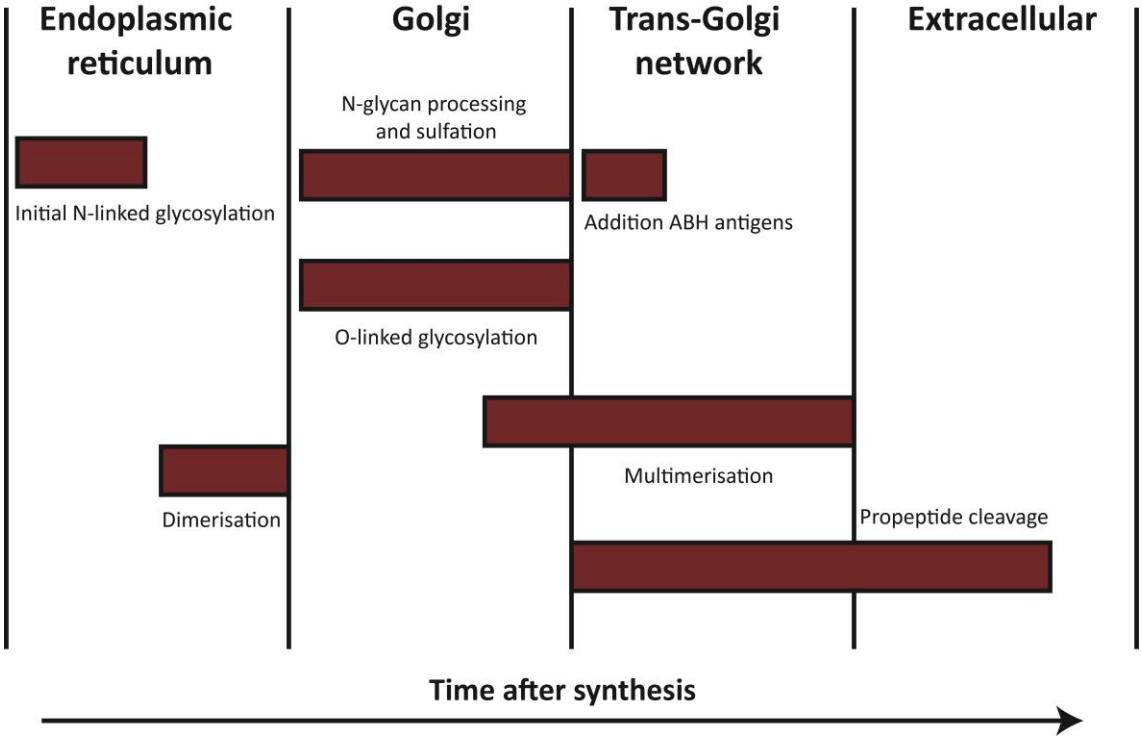


Figure 1.15: Steps in VWF biosynthesis and their subcellular localisation. Following translation, the signal peptide directs VWF to the ER where the signal peptide is cleaved and initial N-linked glycosylation occurs and C-terminal dimers are formed. The molecule passes to the Golgi where there is the addition of O-linked glycans and sulphation of N-terminal N-linked glycans. In the *trans*-Golgi network there is the addition of ABO(H) blood group sugars to N-linked glycans and VWF multimerisation by N-terminal disulphide bonding. The propeptide is cleaved and the molecule is either constitutively secreted or stored in Weibel-Palade bodies.

In the *trans*-Golgi network, further processing of correctly assembled proVWF protomers takes place, including proteolytic removal of the propeptide, tyrosine sulphation of certain N-linked glycans[142,157] and the addition of ten O-linked oligosaccharides to serine or threonine residues (Figure 1.16). Two O-linked glycosylation sites are localised N-terminal to domain A3 and within VWC1 (Thr1679 and Thr2298, respectively) while the remaining eight (Thr1248, Thr1255, Thr1256, Thr1263, Thr1468, Thr1477, Ser1486, Ser1487) are clustered in groups of four at either side of the A1 domain[38,158] (Figure 1.13). 1% of the VWF O-glycome has recently been shown to harbour ABO(H) antigens and these glycans may also participate in determining plasma VWF levels and susceptibility to ADAMTS13 cleavage[158].

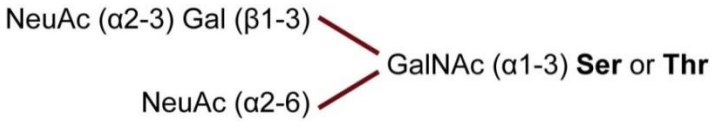


Figure 1.16: Primary structure of the major O-linked glycan. Gal, galactose; NeuAc, *N*-acetylneuraminic (sialic) acid; GalNAc, *N*-acetylgalactosamine.

In the same compartments, multimerisation of proVWF occurs with the formation of additional 'head to head' disulphide bonds through cysteine residues in the D'D3 region (N-terminal), yielding incredibly large polymers[23]. The VWF propeptide (D1D2 region) is cleaved by the enzyme furin, at the ^{P4}Arg-^{P3}x-^{P2}Lys/Arg-^{P1}Arg$^{\downarrow}$ motif (where $^{\downarrow}$ identifies the cleavage site)[159], comprising substrate recognition residues 760 (P4) through 763 (P1), to generate mature VWF. The propeptide is essential to the correct assembly of VWF multimers. Following cleavage, the propeptide remains non-covalently associated with VWF in an interaction promoted by the high calcium concentration and acidic pH of the *trans*-Golgi[23]. VWF multimerisation in the Golgi involves a unique oxidoreductase mechanism that relies on the propeptide to function as an endogenous chaperone that promotes disulphide bond formation specifically under acidic conditions[91]. Deletion of the VWF propeptide prevents multimerisation, and interestingly, the propeptide does not need to be a contiguous part of the primary structure, as isolated propeptide is able to act in *trans* to direct the assembly of disulphide-linked multimers[160]. Further, VWF propeptide cleavage is not necessary for multimerisation, as evidenced by the Arg763Gly mutation, that completely abolishes

proteolytic cleavage but does not significantly affect the multimeric assembly of VWF, in which the multimers are composed of proVWF subunits[161,162].

The Golgi is a hostile environment for disulphide bond formation, as the low pH inhibits deprotonation of the cysteine thiol groups. Protein disulphide formation proceeds predominantly in the lumen of the ER and this is the site of polymerisation of most oligomeric proteins[93,163]. Nevertheless, VWF multimerisation is optimal at pH 5.8 and in fact, an acidic *trans*-Golgi environment (pH 6.2) may be necessary because raising the pH with ammonium chloride or chloroquine inhibits multimerisation[40,91,164]. The decrease in pH between the ER and Golgi would facilitate the reversible protonation of a pH-sensing His residue. Histidine has a pK_a of ~6.0, a value suitable for detecting the difference in pH between the ER (pH 7.4) and *trans*-Golgi (pH 6.2)[165], and is likely to perform this function during the assembly of VWF multimers. Several proteins employ His residues as pH sensors to regulate their function upon transport to the Golgi. His69 in the autoinhibitory propeptide of furin becomes protonated in the *trans*-Golgi, after which the propeptide dissociates and furin can cleave proVWF. There are at least two evolutionary conserved His residues in domain VWD2 (His395 and His460) that are critical for pH-dependant multimer assembly in the Golgi, as demonstrated by the profound defect in intersubunit disulphide bond formation upon mutation to alanine[40]. This is consistent with the observation that the VWF propeptide forms non-covalent homodimers, which bind tightly to multimeric VWF at pH 6.4 but dissociate at neutral pH. This would facilitate the juxtaposition of two VWF D3 regions, and in this manner correctly orientate dimeric VWF for multimerisation upon arrival at the Golgi[166].

Intriguingly, to overcome the impediment of an acidic environment, VWF employs its own propeptide as an oxidoreductase[131]. The VWD1 and VWD2 domains contain CxxC sequences, Cys159-Gly-Leu-Cys162 and Cys521-Gly-Leu-Cys524, which is the signature motif of the functional sites of the thiol:disulphide oxidoreductase family. This conserved motif is characterised by a pair of vicinal cysteine residues that shuffle between the disulphide and dithiol form[163]. When spacing of this sequence is altered by insertion of a glycine residue, into either VWD1 or VWD2, thus changing it to CxGxC, dimers were secreted without multimer assembly, therefore the propeptide catalyses

multimerisation by its intrinsic pH-dependant disulphide isomerase activity[93]. The role of the CGLC motif in domain VWD3, at the N-terminus of the mature protein, is unknown. Mutation of the corresponding VWD3 motif in mucins permits multimer formation in the presence of monensin, suggesting that an intact VWD3 CGLC motif acts to prevent multimerisation in the non-acidic ER, *cis*- and *medial*-Golgi compartments[89], the mechanism by which this may be accomplished is not clear.

The thiol:disulphide oxidoreductase model for VWF propeptide function predicts that rearrangements would proceed through a covalent disulphide-linked intermediate between a catalytic Cys residue on the oxidoreductase and a Cys residue to be modified on the substrate. Indeed, it has been shown that the propeptide and the D3 region of VWF subunit form an intra-chain disulphide-bonded intermediate in the ER that rearranges in the Golgi to yield VWF multimers, linked by inter-dimer disulphide bonds[92] (Figure 1.17). The D'D3 region of VWF was sufficient to form this intermediate with the propeptide, thereby confining the mechanistically important thiols participating in multimerisation to the N-terminus. These cysteine residues have been localised to Cys1099 and Cys1142 in domains C83 and TIL3, respectively[43].

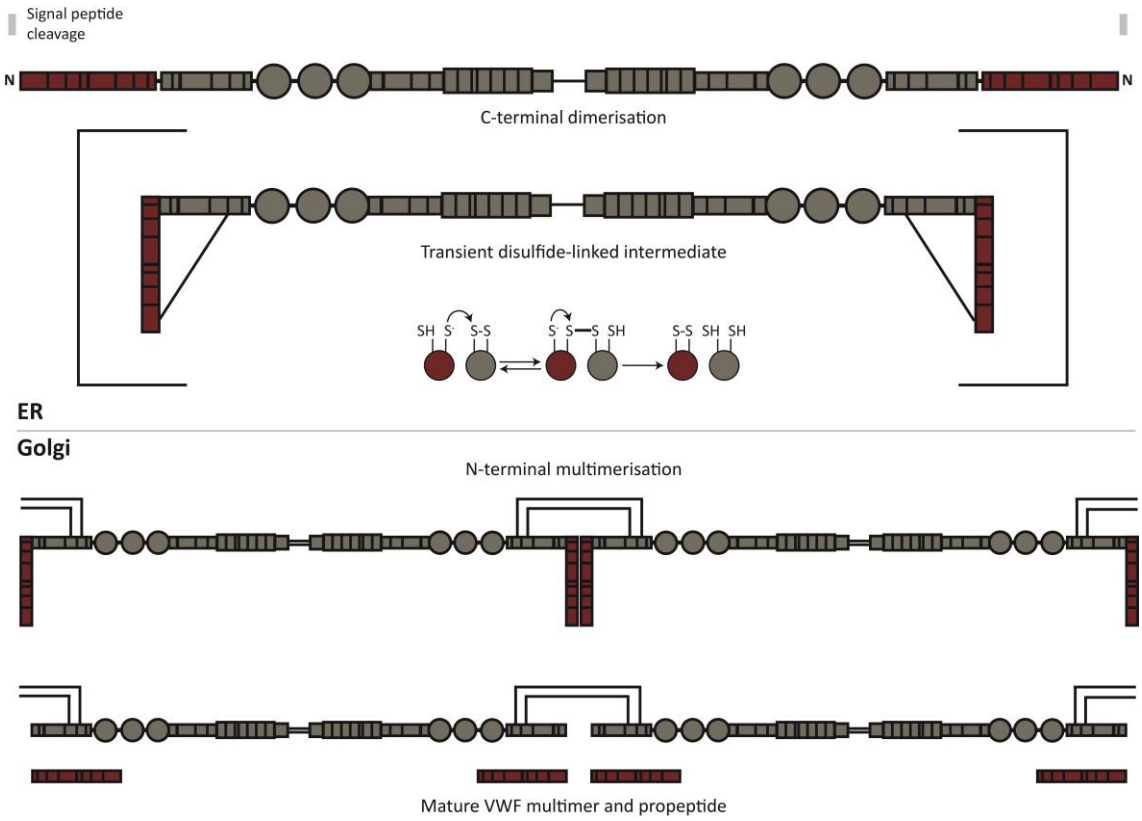


Figure 1.17: Biosynthesis of VWF. The VWF propeptide (D1D2 region) is shown in red and the mature protein is shown in grey. VWF is synthesised in the ER as a 2813 aa prepro-protein. In the ER VWF forms tail-to-tail dimers between C-terminal cysteine knot (CK) domains. Disulphide bonds are represented by black lines. Later in the ER, a transient disulphide-linked intermediate is predicted to form between the propeptide and the D3 region of VWF, the mechanism of which is represented by balls. In the Golgi the disulphide-linked intermediate rearranges to yield head-to-head linked multimers. Disulphide bonds form between the D3 regions. In the *trans*-Golgi furin cleaves the propeptide.

1.6.3 Storage and secretion

VWF is partitioned between two pathways in the Golgi of endothelial cells. The majority, up to 95%, is secreted constitutively towards both the lumen (apical secretion) and the subendothelial matrix (basolateral secretion)[167]. The remainder is directed to storage by the propeptide, for later inducible release, into Weibel-Palade bodies (WPBs). VWF multimers released through the constitutive pathway are smaller, and hence, less thrombogenic than those stored in WPBs, which are unique to vascular endothelial cells and enriched in ultra large VWF (ULVWF) forms. VWF produced by megakaryocytes is stored in α-granules that are later partitioned into platelets. VWF

circulates in plasma at a concentration of ~10 μg/mL, has a half-life of approximately twelve hours and is cleared from the circulation by hepatic and splenic macrophages[168].

VWF undergoes extraordinary structural changes during its synthesis and storage and one of the enduring mysteries of VWF biology is how incredibly large multimers are condensed and packaged into the cigar-shaped WPBs. The details by which VWF achieves this remarkable feat are beginning to emerge. VWF multimers are stored as paracrystalline helical tubules with a precise parallel alignment[169] that extends the entire length of the WPB, 1-5 μm[30]. In contrast to DNA, with two molecules per helix, the VWF helices are one start, that is, there is one VWF multimeric molecule per helix[166], hence a single helical VWF multimer may extend up to 5 μm. The tubules consist of a right-handed helix with 4.2 repeating units (dimers) per turn and a pitch of 12 nm[166,169]. Indicating that every 1 μm of VWF tubule length, might correspond to a disulphide-linked multimer of 700 subunits (350 dimers), having a mass of 175 million Da and an extended length of 49 μm (average extension per monomer of 70 nm[140]).

The organisation of VWF multimers into helical tubules nucleates in the *trans*-Golgi and relies solely on the propeptide and the D'D3 region, requiring Ca^{2+} and a low pH as found in the *trans*-Golgi and WPB environments. The D1D2 propeptide and the D'D3 regions form the core of the helix, with the remainder of the VWF subunits (domains A1 to CK) extend from the helix core and may interdigitate to account for the regular spacing seen between VWF tubules in a WPB[166,169,170]. The A2 to CK domains of each subunit of a dimer assemble into a stem-like extended conformation (termed the dimeric bouquet) at acidic pH representative of the *trans*-Golgi (pH 6.2) and WPB environments (pH 5.45)[140,171]. In each dimer pair, these C-terminal domains are brought into very close association, and in conjunction with the propeptide-VWF dimer association, they ensure co-linear multimerisation of VWF protomers and are thus, crucial for the arrangement of VWF into tubules. This arrangement facilitates disulphide bond formation between the D3 region of VWF dimers in neighbouring positions in the assembling helical tubule, rather than within VWF dimers, between monomers on different tubules or in non-neighbouring positions in the same tubule. Covalent addition of a new VWF dimer to the end of the growing helix, catalysed by

the propeptide, provides a strategy for packaging the long VWF multimers without entanglement[169]. This mechanism, unbelievably, enables a 50-fold compaction in the length of tubules compared with extended secreted VWF[166], and compaction is likely to be assisted by the reduction in pH in mature WPBs, which approaches the isoelectric point of VWF[171]. Furthermore, compaction is particularly extreme at the C-terminal, 51 nm long, dimeric bouquet region of VWF that extends out from the core of the helix[140,172], as the 25 nm diameter tubules pack with a centre-to-centre spacing of 28.4 nm[169]. Tubule formation enables orderly unravelling of VWF from the ends of helices when secreted into plasma[140].

At the *trans*-Golgi level, disorganised tubules are packed into vesicles that bud off, leading to the formation of immature WPBs. Those containing lower molecular weight VWF tubules are destined for constitutive release[173]. During maturation, WPBs undergo homotypic fusion thereby increasing their VWF tubule content and subsequently a dense paracrystalline core is formed by the alignment and condensation of VWF tubules. An event characterised by a reduction in the WPB diameter from 222 nm to 151 nm[169,170,173]. Once they have entered the regulated secretory pathway, mature WPBs lie dormant in the endothelial cell awaiting a signal to undergo exocytosis and release VWF into circulation. A myriad secretagogues can induce the release of VWF from WPBs and these are divided into two groups: Ca^{2+}-raising agonists and cyclic AMP (cAMP) raising agonists[174]. 1-deamino-8-D-arginine vasopressin (DDAVP), a pharmaceutical analogue of vasopressin (antidiuretic hormone), binds to endothelial V2 receptors leading to production of cAMP and activation of protein kinase A, consequently stimulating WPB exocytosis[175]. This results in a transient increase in VWF and FVIII plasma levels and has led to the widespread use of DDAVP in the treatment of patients with VWD and mild haemophilia A[175].

The first step in secretion is the formation of a narrow constriction between the endothelial cell plasma membrane and the WPB. This results in WPB swelling, as a consequence of the rise in pH, where alkalinisation rapidly propagates from one end of the WPB to the other, and partial hydration of the granule core[171]. The WPB morphology change is associated with loss of alignment in the VWF paracrystal[169]. Secretion proceeds with the rapid unfurling of VWF helices[172] into long strings at the

luminal endothelial cell surface. Plasma VWF is disorganised and globular in solution and when extended in high shear flow can have lengths of up to 4 μm[139]. In dramatic comparison, following secretion from endothelial cells, ULVWF strings often reach extraordinary lengths of 100-1000 μm and they remain anchored to the surface of the endothelial cell[176,177]. As each monomer in an extended configuration spans ~70 nm[140,172], an unravelled 5 μm VWF helical tubule would be expected to measure ~245 μm. Thus, the exceptionally long VWF strings, observed at lengths of up to 1 mm, are thought to arise following WPB secretion through end-to-end self-association, as well as lateral staggered association that stacks the multimers into thick fibrillar structures in shear flow. The mechanism appears to involve cysteine thiol:disulphide exchange of unpaired thiols in mature VWF[44–46], of which the VWC3 Cys2431-Cys2453 bond is believed to be involved[178] and the process modulated by ADAMTS13[179].

1.6.4 Remodelling VWF multimer distribution

Whereas globular plasma VWF is functionally quiescent in the absence of shear, due to shielding of binding sites, ULVWF released by the inducible pathway at sites of vascular injury is highly thrombogenic and hyperactive in its capacity to bind platelets and subendothelial matrix[128]. This difference relates to both the multivalent nature of VWF binding, with multiplicity of available collagen and platelet binding sites, and the propensity of ULVWF to adopt an elongated, active conformation in the normal circulating shear forces of flowing blood; as the forces acting on proteins in shear flow scale with protein size and length[180–182]. Consequently, their release may result in spontaneous platelet adhesion, aggregation and systemic thrombosis[181]. This is the pathophysiology underlying the catastrophic arterial thrombotic disorder, thrombotic thrombocytopenic purpura (TTP)[183], that is characterised by extensive microvascular occlusion with thrombi composed almost entirely of VWF and platelets. The clinical consequence of which, is microangiopathic haemolytic anaemia and thrombocytopenia leading to renal failure, neurologic dysfunction and fever as signs of end organ damage. Without treatment, TTP undertakes a rapid course of deterioration and death, in most cases[184,185]. TTP is caused by severe deficiency in ADAMTS13, which

is responsible for the processing of ULVWF to less reactive plasma forms[186] and is the only known mechanism that governs VWF multimer size. Hereditary TTP is caused by homozygous or compound heterozygous mutations in the *ADAMTS13* gene[186], although ADAMTS13 deficiency more commonly arises through acquired TTP involving the formation of inhibitory anti-ADAMTS13 autoantibodies[184].

To prevent this disastrous consequence, ULVWF is subject to rapid but limited proteolysis by ADAMTS13 through cleavage at the Tyr1605-Met1606 peptide bond in the A2 domain[72]. Within two hours, following WPB exocytosis from endothelial cells, ULVWF is remodelled into smaller, haemostatically active, but no longer prothrombotic forms, of variable size that are characteristic of the circulating pool of globular VWF[187] (Figure 1.18). ADAMTS13 is constitutively active in plasma, unusually therefore, VWF proteolysis is regulated by conformational change[188,189]. The ADAMTS13 cleavage site is normally hidden within the core of the folded A2 domain, and when plasma VWF is circulating in its inactive globular conformation it is resistant to proteolysis[190]. However, shear-induced unfolding and elongation not only exposes the scissile bond but also cryptic exosites for ADAMTS13[71,182]. ULVWF freshly released from endothelial cells spontaneously adopts an elongated configuration enabling ADAMTS13 to both access and cleave the exposed A2 domains[180,182]. Thrombospondin-1 (TSP-1) also appears to regulate cleavage of VWF strings by ADAMTS13. The inhibitory effect of TSP-1 may be explained by competition between TSP-1 and ADAMTS13 for the same binding sites in VWF A2 and A3 domains[191].

Conversely, excess ADAMTS13 mediated proteolysis would severely impair the haemostatic function of aberrantly short VWF multimers, leading to bleeding, as found in patients with type 2A (IIA) VWD[192] who harbour mutations that destabilise the A2 domain - resulting in enhanced cleavage[193].

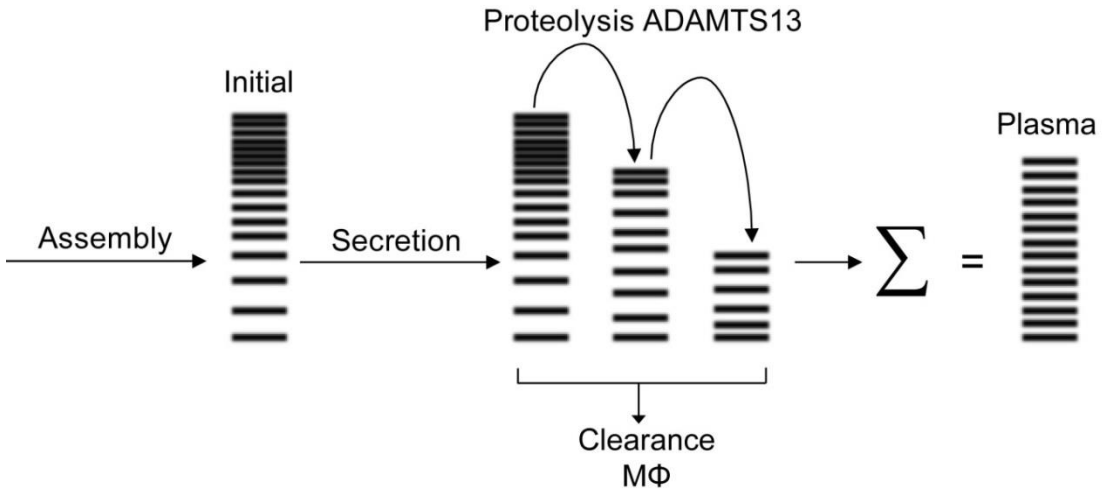


Figure 1.18: ADAMTS13 remodelling of VWF multimers. VWF multimers are assembled in the *trans*-Golgi as an initial distribution (Initial) and secreted into the blood constitutively or from WPBs. This multimer distribution is modified by rate of ADAMTS13 proteolysis and clearance by hepatic and splenic macrophages (Φ). The distribution of plasma VWF reflects the sum of these effects.

1.7 VWF in haemostasis and structure-function relationships

The three significant interactions of VWF in haemostasis are with collagens, platelets and procoagulant FVIII, to which the multidomain, multimeric configuration of VWF is fundamental. Much of the function of VWF has been revealed, however detailed insight into the molecular structure that enables VWF to orchestrate haemostatic processes, in particular FVIII stabilisation in plasma, is lacking. In addition, high resolution structure determinations of the topologically complex disulphide-bonded VWF domains have not been possible, with the exception of the cassette of A domains containing only a pair of cysteines each, for which there are X-ray crystal structures[62,63,190,194–197].

1.7.1 Shear and VWF activation

A remarkable feature of VWF is that it is the key shear-sensing protein in haemostasis, and is especially important in arterial bleeding where shear is high[23]. ULVWF and plasma VWF adopt different conformations that make the former constitutively active and the latter inactive, but can be activated by shear induced elongation[138]. Once

released into circulation, the overall shape of VWF is dynamic and dependent on the type of flow encountered (Figure 1.19).

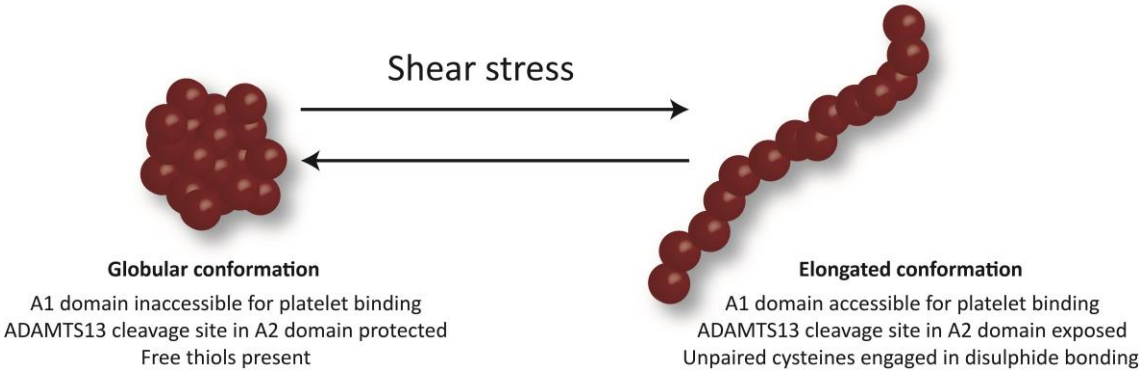


Figure 1.19: Events associated with shear-induced conformational changes in VWF. VWF circulates in a globular conformation in which the platelet GPIbα binding site in the A1 domain and ADAMTS13 cleavage sites are inaccessible. Both parameters protect VWF from premature platelet interactions and ADAMTS13-mediated degradation. Exposure to high shear flow unfolds the protein, allowing platelet binding and ADAMTS13 cleavage.

As blood flows through a vessel, velocity is maximal in the centre and falls to zero at the wall. Conversely, the gradient of velocity, or shear rate, γ (expressed in units of cm/s per cm or s^{-1}), is maximal at the vessel wall and zero at the centre. Part of the effect of shear is to increase the transport of platelets to the vessel wall, but more importantly to increase the reactivity of VWF by increasing the number of accessible binding sites[23]. The circulatory system is exposed to an amazingly wide span of shear rates (Table 1.1).

Table 1.1: Shear rates within the human vasculature[198].

Vessel	Diameter (mm)	Shear Rate γ (s^{-1})
Ascending aorta	23.0-45.0	50-300
Femoral artery	5.0	300
Common carotid	5.9	250
Internal carotid	6.1	220
Left main coronary	4.0	460
Right coronary	3.4	440
Small arteries	0.3	1500
Arterioles	0.03	1900
Large veins	5.0-10.0	200
Inferior vena cava	20.0	40-60

The hydrodynamic drag that opposes platelet adhesion and aggregation on the luminal surface exposed to rapidly flowing blood, increases with shear rate. Consequently, its effects on VWF and thrombus formation are more relevant in arteries, where platelet adhesion is dependent on VWF, than in veins[2,35]. Above a critical shear rate of ~1000 s^{-1}, VWF in free flow is overcome by hydrodynamic drag and the molecules elongate, becoming functionally multivalent[172,199–201]. *In vivo*, shear rates in this range occur in small arterioles of 10-50 μm diameter, where the shear rates are estimated to vary between 470 s^{-1} and 4,700 s^{-1} (median, 1,900 s^{-1}). Above atherosclerotic plaques, in partially occluded arteries, shear rates of up to 11,000 s^{-1} have been recorded[23,202]. VWF acts as a molecular bridge, and is central to the initiation of platelet adhesion and sustaining platelet aggregation under conditions of elevated shear[2]. Thus, the two-fold consequence of elevated shear, causing a near vessel wall platelet excess and activating VWF are of utmost strategic importance to VWF in haemostasis (Figure 1.20). Conversely, that same capacity to promote platelet adhesion in high shear rates, conditions found in stenosed arteries, also makes VWF a crucial component of atherothrombosis in cardiovascular disease and ischaemic stroke[203–206].

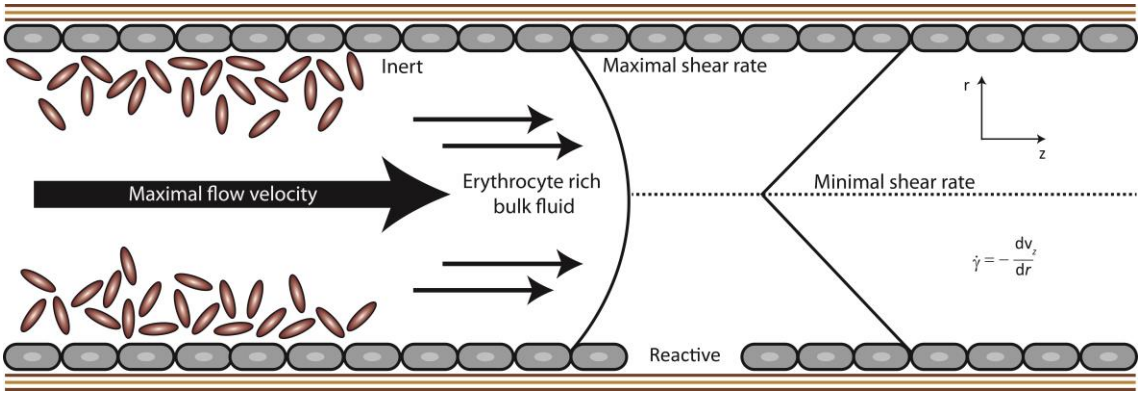


Figure 1.20: Blood flow in a vessel. Normal endothelial cells are nonreactive for platelets, but exposed subendothelial structures such as collagen, induce rapid platelet adhesion and aggregation. Blood flow is laminar, with each layer moving at a different velocity. The laminae near the centre of the vessel have greater velocity than those near the wall (depicted by arrows). The shear rate is the rate of change of velocity with respect to distance measured perpendicularly to the direction of flow. The negative sign indicates that the gradient is defined from the centre (where velocity is maximal) to the wall (where velocity is minimal). Because of these rheological properties, erythrocytes occupy the greatest part of the lumen, pushing leucocytes and platelets to the sides creating a near-wall platelet excess[2].

1.7.2 VWF A1 and A3 domains – collagen binding

The subendothelial matrix, of which fibrillar collagens are a major component, is exposed when the continuity of the vascular tree is interrupted by traumatic injury to tissues. In the wounded vessel, VWF rapidly binds to collagens type I and III in the deeper layers of the vessel (tunica media and tunica adventitia), as well as microfibrillar collagen type VI in the subendothelium[2]. This is substantially contributed to by subendothelial VWF, as a pool distinct from plasma VWF[207,208]. The transition from soluble plasma VWF to immobilised VWF marks the initiating event in thrombogenesis. The VWF A1 domain preferentially interacts with collagen type VI, whereas VWF binds to collagens type I and III through the A3 domain. The A1 domain also contains the binding site for platelet GPIbα of the GPIb-IX-V complex. The latter, A3 mediated collagen interaction is apparently both necessary and sufficient to support VWF binding to collagens[57,209,210]. It is of note however, that the extensive investigation of the interaction between VWF A1 and A3 domains and fibrillar collagens has been primarily undertaken in static conditions using recombinant soluble VWF with deleted A3, which demonstrates antithrombotic activity. In contrast stands the *in vivo* evidence that mutations in domain A3 (Ser1731Thr, Ile1741The, Gln1743H, Trp1745Cys, Gln1762Arg, Ser1783Ala, His1786Asp) inhibiting collagen binding (type 2M VWD; classified by decreased VWF-dependant platelet adhesion) are still capable of sustaining haemostasis[211–214]. Therefore, fluid dynamic conditions and mechanical forces may modulate these interactions, with a complex interplay between domains A1 and A3 in effecting immobilisation of VWF, which is contingent on the nature of the lesion[2]. Such that binding of domain A1 to collagen type VI may be particularly relevant in A3 defective mutants, and may account for the lack of severe bleeding in these patients[214].

VWF is also able to interact with the subendothelial matrix independently of collagen. The A1 domain contains a heparin-binding site, Tyr1328-Ala1350 that may reflect the ability of VWF to interact with matrix proteoglycans that contain sulphated carbohydrates[2,215]. In addition, VWF binds to sulphated glycosphingolipids (sulphatides) that are present on various cellular membranes and are likely to serve an accessory role in promoting localisation on wounded tissue[78]. Sulphatides are

constituents of erythrocytes, but are not normally exposed; however, in sickle cell anaemia reversible sickling and oxidant damage to the membrane could expose these molecules to the surface. The subsequent interaction between VWF and exposed sulphatide may cause periodic vaso-occlusion in the affected patient, the severest manifestation of sickle cell anaemia[2,182].

As the principal activator of thrombosis, a mechanistic understanding of the VWF-collagen interaction is vital. The structure of the A3 domain has been determined by X-ray crystallography[195,196]. The VWF A domains are the prototype for the von Willebrand factor type A domain (VWA) superfamily, that are found in many proteins including collagens, complement proteins and integrins[66]. The A3 domain adopts the dinucleotide-binding fold, consisting of a central hydrophobic parallel β-sheet containing six strands that is enclaved by seven α-helices. A single disulphide bridge between residues Cys1688 and Cys1872 connects the N- and C-terminal regions and denotes the domain boundaries. The VWA domains contain a metal ion dependant adhesion site (MIDAS) motif at the top of the β-sheet that is critical for its adhesive function, however, the VWF A3 domain is notable for its absence of the MIDAS motif, indicating that metal ions are not required for collagen binding[195,196].

The collagen binding site was localised by transferred cross-saturation, a nuclear magnetic resonance (NMR) method enabling rigorous identification of contact residues in a protein complex[67] and it was subsequently structurally characterised by crystallisation of the VWF A3-collagen complex[197]. Collagen is formed of three tightly interwoven polypeptide chains (α-chains) in a right-handed triple helix with a one residue stagger. These α-chains have the repeating sequence G-x-x', where x is frequently proline (P) and x' is frequently 4-hydroxyproline (O). The VWF binding site on collagen III is located on residues 572-580 with the sequence: RGQOGVMGF[216]. In collagen III, these residues form a unique VWF interaction surface to which all three α-chains contribute. The A3 domain binds collagen at its front face, across strand β3 and helices α2 and α3; the binding site can be divided into three regions (Figure 1.21). First, a small hydrophobic patch on A3 interacts with the N-terminus of the collagen III recognition sequence. Second, interacting with the central portion of the collagen sequence is an A3 acidic region. Third, the hydrophobic C-terminus of the collagen

recognition sequence interacts with a large hydrophobic patch on A3 including a shallow hydrophobic pocket that binds collagen sequence Phe (F-pocket). Comparison of free and collagen-bound VWF A3 domains shows that collagen binding occurs through the conformational selection model of ligand binding. In this model, the $\alpha 3\beta 4$ loop of domain A3 adopts a conformation that facilitates formation of the hydrophobic F-pocket, which is stabilised by a network of hydrogen bonds between His1786, Ser1783 and the Met residue (M) in the collagen binding sequence. This hydrogen-bonding network is crucial to collagen binding as evidenced by the type 2M VWD mutant Ser1783Ala that completely abolishes VWF binding to collagen[212].

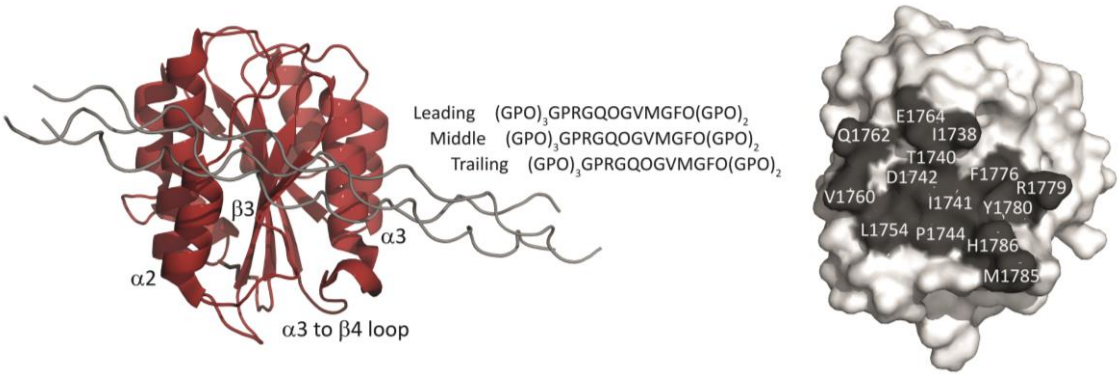


Figure 1.21: Structure of the VWF A3-collagen peptide complex. The left panel shows the VWF A3 domain in red by ribbon representation, with the $\alpha 2$, $\beta 3$, $\alpha 3$ secondary structure elements labelled. The $\alpha 3$ to $\beta 4$ loop is indicated. The triple helical collagen peptide in complex with A3 is shown in grey (pdb: 4DMU)[197]. The one residue stagger in collagen results in a leading, middle and trailing chain. The core VWF binding sequence on collagen is shown and it is flanked by glycine-proline-hydroxyproline (GPO) repeats. The right panel shows the surface representation of VWF A3 domain in white and the residues interacting with collagen are highlighted in grey. The residues are denoted by single-letter amino acid codes.

Platelets have no measurable interaction with soluble VWF and the obligatory perquisite to platelet adhesion is VWF immobilisation onto collagen. Inhibition of VWF binding to collagens by targeting the nonapeptide collagen sequence may be a useful therapeutic strategy. *In vivo* studies have demonstrated that targeting VWF-collagen binding can prevent platelet thrombosis on an otherwise highly thrombogenic injured artery. This antithrombotic approach may provide a valuable alternative to the antiplatelet therapies in current clinical practice[217,218].

1.7.3 VWF A1 domain - platelet GPIbα tethering

In a blood vessel with normal cell counts, erythrocytes represent the greatest cellular mass, and as a consequence of their erratic motion and lateral migration, they occupy the greatest part of the lumen. This pushes leucocytes and platelets to the vessel sides, creating a near-wall platelet excess, where platelets are flawlessly placed to facilitate adhesive interaction with immobilised VWF (Figure 1.20). The clinical observation that bleeding, particularly spontaneous, is usually not of consequence until the platelet count drops below ~10,000/μL, is explained by the near-wall platelet excess. While formation of the large platelet aggregates required to seal post-traumatic wounds in vessels may be affected by even a relatively small reduction in platelet number, the platelet monolayer that maintains vascular permeability (and thus, prevents spontaneous haemorrhage) is largely undepleted until the platelet count drops significantly (1% of normal). Furthermore, since the near-wall platelet excess is consequent to the presence of erythrocyte mass, it is then clear why anaemia may predispose to excessive bleeding[2].

Above a threshold shear rate of 500-800 s^{-1}, only the interaction between immobilised VWF domain A1 and platelet GPIbα has a sufficiently fast on-rate to initiate platelet adhesion[2,219], as the intrinsic association rate of the platelet receptor with its ligand must be faster than the relative velocity at which the molecules move relative to each other. Remarkably, VWF will be most extended in just those environments where its haemostatic functions are needed most. Shear rate is greatest at the vessel wall, and when immobilised onto collagen, VWF cannot tumble and thus, the hydrodynamic forces acting on vessel-bound VWF are greater than on plasma VWF, and these forces are purely elongational[172,220]. This phenomenon serves to induce a conformational change in VWF that exposes A1 domain sequences for initial interaction with platelet GPIbα receptors[221,222]. An additional level of regulation utilises the N-terminus flanking region of the A1 domain to stabilise the quaternary association of the A1-A2-A3 domains, which in turn regulates VWF A1-GPIbα binding[223,224]. Circulating VWF propeptide has also been implicated in the partial regulation of VWF A1-platelet GPIbα complex formation[225].

Platelets bind specifically to immobilised VWF in a two-stage process that involves sequentially, GPIbα and integrin αIIbβ3. Initial tethering is mediated by GPIbα-VWF A1 domain bonds that form extremely rapidly but have a limited lifetime and hence, cannot support stable adhesion[219]. Rather, these bonds support a slow rolling movement along the adhesive VWF surface in the direction of blood flow, and platelet translocation is associated with platelet activation. Interestingly, the bond between GPIbα and VWF has two states; distinct, flexed low affinity and extended, high affinity states[226]. Increasing shear exponentiates the GPIbα-VWF bond lifetime causing platelet rolling velocity to decrease and facilitates platelet activation and subsequent integrin mediated attachment[219,226,227]. The mechanochemical force-resistant properties of the GPIbα-VWF bond[226] requires high shear rates to achieve functionally efficient lifetimes, and as such, a property that may prevent VWF A1 domains from agglutinating platelets in the circulation[227]. The inherently stickier ULVWF released by stimulated endothelial cells may enhance the efficiency of the process, as these molecules spontaneously form high strength bonds with GPIbα[181].

The unusual variant, type 2B VWD, is due to gain-of-function mutations in the VWF A1 domain that lead to constitutive VWF A1-GPIbα complex formation, and can explain the ability of type 2B VWF to agglutinate platelets in circulating blood[2]. Type 2B mutations in the VWF A1 domain, eliminate the shear threshold requirement for platelet rolling, and result in an enhanced A1-GPIbα association rate with relatively long bond lifetimes, even when exposed to low shear rates[172,227,228]. This is consistent with the significantly slower rolling, with near zero velocity of platelets interacting with immobilised type 2B VWF A1 domains[2,228]. Paradoxically, the enhanced GPIbα-VWF A1 bonding in these patients is associated with excessive bleeding. This is due to depletion of the large, haemostatically most active VWF multimers from plasma, relapsing thrombocytopenia due to platelet consumption, and the occupancy of GPIbα receptors that interferes with the tethering of platelets to VWF at the site of vascular injury[229]. This phenotype is additionally contributed to by excessive ADAMTS13-mediated proteolysis[229,230]. Conversely, the Bernard-Soulier syndrome and type 2M VWD result in defective VWF A1-GPIbα binding secondary to loss-of-function mutations in the GPIb-IX-V complex and VWF A1, respectively[231].

Crystal structures of GPIbα-VWF A1 complex[62,63] have provided important insight into the molecular details of their interaction. Crystallisation of the complex, in one instance, required use of gain-of-function mutations; type 2B mutant Arg1306Gln in A1 and platelet-type VWD mutant Met239Val in GPIbα[62]. As expected the overall VWF A1 domain structure is very similar to the A3 domain with a disulphide bridge linking the N- and C-proximal sequences (Cys1272-Cys1458). GPIbα is the central component of the unique GPIb-IX-V receptor complex consisting of transmembrane glycoproteins Ibα, Iβ, IX and V, each of which, is a member of the leucine-rich repeat protein superfamily. GPIbα anchors the complex to the platelet cytoskeleton and harbours the VWF binding site in its 282 N-terminal residues. The VWF binding site on GPIbα is formed of the N-terminal eight tandem leucine-rich repeats that are exposed well above the platelet surface by a connecting ~45 nm long highly O-glycosylated stalk, which is anchored to the transmembrane domain, and followed by the cytoplasmic tail[232].

Crystallographic studies reveal that the leucine-rich repeats wrap around one side of the VWF A1 domain, but only the N-terminal β-finger region and the C-terminal β-switch region make contact with A1[62,63] (Figure 1.22).

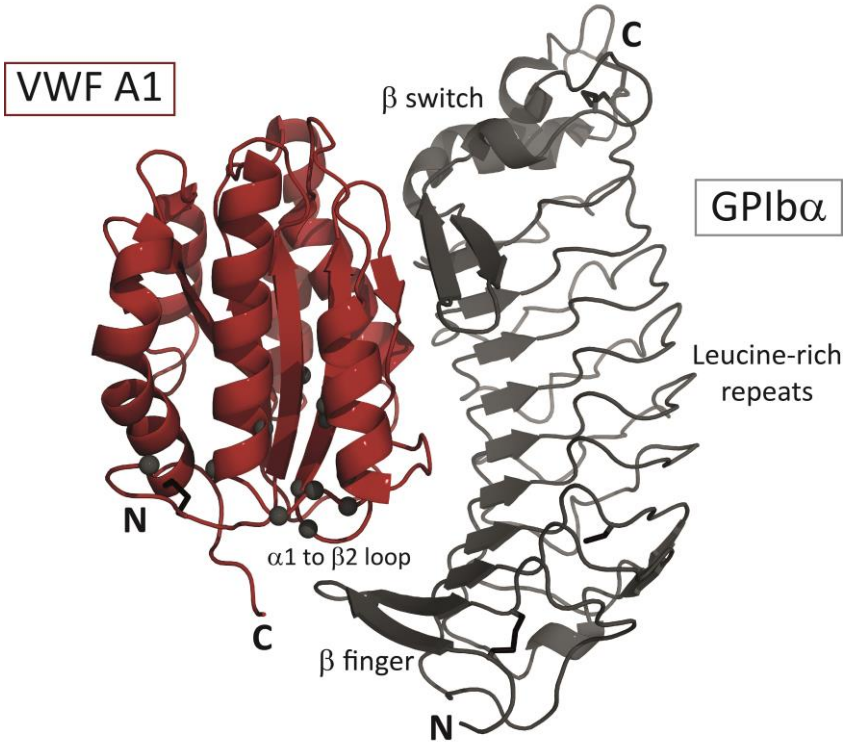


Figure 1.22: Structure of the complex between VWF A1 domain and platelet GPIbα A ribbon representation of the complex is shown (pdb:1M10). VWF A1 domain is coloured red and the N-terminal VWF binding region of GPIbα is coloured grey. This region is formed, centrally, of eight short leucine-rich repeats. The N-terminus of GPIbα is formed of a β-hairpin, termed the β-finger. The C-terminus of GPIbα contains a disordered loop termed β-switch. The positions of type 2B VWD mutations are shown as carbon-α spheres in black.

In unliganded A1 domain however, the lower region of GPIbα contact is shielded by the N- and C-terminal arms of A1, creating an interface that is stabilised by salt-bridges and hydrophobic contacts, thereby making the β-finger binding site on A1 inaccessible to platelets in plasma[194]. Type 2B VWD mutations are clustered at the lower face of domain A1, in a single disulphide loop between residues Cys1272 and Cys1458, encompassing the GPIbα β-finger interaction site, and map to buried positions in domain A1 [62,63,194] (Figure 1.22). These mutations are likely to destabilise the interface created by the folded conformation of the A1 N- and C-terminal arms, which protects the β-finger interaction site. In the crystal structure of the complex, indeed a displacement of the A1 N- and C-termini away from the interaction site was observed[62]. Furthermore, the α1-to-β1 loop acts as a conformational switch, that in unliganded A1 adopts a conformation that creates a steric clash with GPIbα and would therefore, be incompatible with platelet binding. In wild-type A1 complexed with

GPIbα the α1-to-β1 loop is more open and shifted away from GPIbα allowing faster dissociation, whereas the loop adopts a more closed and extended conformation that is shifted towards GPIbα in type 2B mutants, which would stabilise a tighter association with GPIbα[63] (Figure 1.23). These conformational transitions illustrate the manner in which the GPIbα-A1 interaction may be modulated by remote type 2B amino acid substitutions, through allosteric changes at the binding face and conformational rearrangements induced by shear, to facilitate VWF-platelet tethering[63].

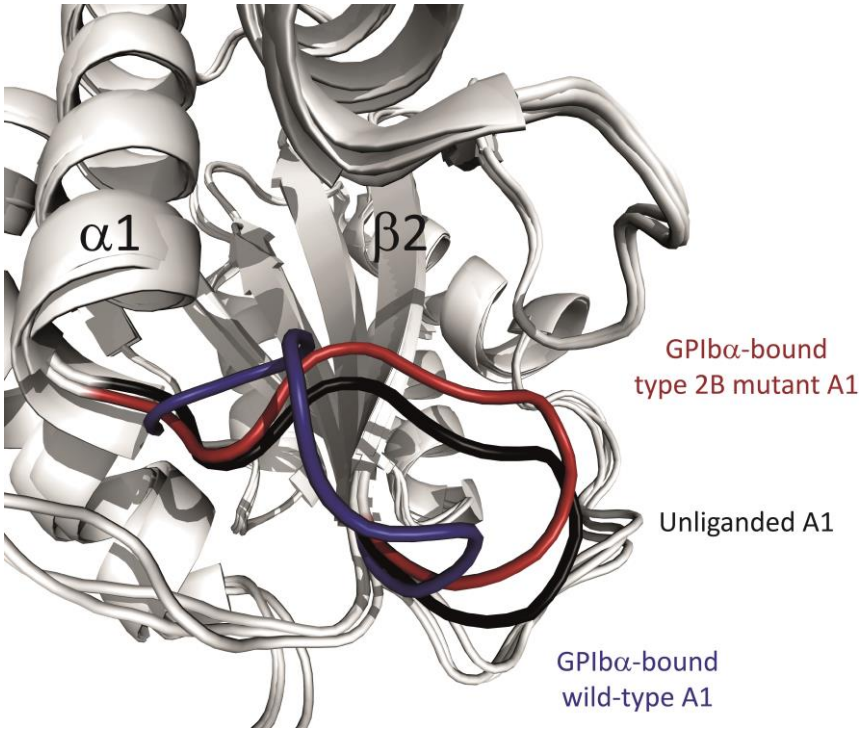


Figure 1.23: A1 α1-to-β2 loop: a conformational switch in GPIbα binding. Superposition of unliganded A1 (pdb: 1AUQ)[194], wild-type A1-GPIbα complex (pdb: 1SQ0)[63] and mutant Arg1306Gln A1-Met239Val GPIbα complex (pdb: 1M10)[62]. Unliganded A1 (black) adopts an intermediate conformation, between the closed conformation in the mutant complex (red) and the open α1 to β2 loop conformation in the wild-type complex (blue).

1.7.4 VWC4 domain – platelet integrin αIIbβ3 aggregation

Following the initial platelet tethering, intracellular signalling through the GPIb-IX-V complex induces pulses of increased cytoplasmic Ca^{2+} concentration, indicating platelet activation[2]. Plasma VWF is then able to bind to the surface of adherent and activated platelets. Activated platelet integrin αIIbβ3 receptors irreversibly bind to the RGD sequence of VWC4 domains leading to stabilisation of the initial GPIbα-VWF A1 bond and definitive platelet arrest. This new layer of VWF forms the substrate onto which more flowing platelets attach through their GPIbα receptors. Accrual of additional platelets in successive layers is secured by the binding and activation of newly recruited platelets. A process that continues until the thrombus mass grows sufficiently to stem bleeding (haemostasis) or in pathological conditions, until the vascular lumen is occluded (thrombosis)[233]. In addition to VWF binding, activated αIIbβ3 receptors also bind to fibrinogen, an interaction that strengthens the adhesion and stability of aggregating platelets[234]. Congenital absence the αIIbβ3 as in Glanzmann's thrombasthenia results in a severe bleeding phenotype, thereby underscoring its importance as an adhesive receptor in platelet aggregation[235]. Missense mutations in VWF VWC4 domain are associated with type 1 VWD (type 1, partial quantitative deficiency representing ~80% of VWD cases) and there are no reported mutations affecting the RGD tripeptide sequence (aa 2507-2509)[236]. Structural characterisation of these mutations or detailed insight into the mechanism of VWC4-αIIbβ3 binding is as yet not possible, due to the paucity of structural data for this region of VWF.

Vascular damage-induced exposure of subendothelial tissue factor (extrinsic pathway) and co-localised FVIII (intrinsic pathway) by its carrier protein VWF, initiate secondary haemostasis at the site of the growing thrombus. Propagation of the coagulation cascade occurs by localised enzymatic complexes assembled on the activated-platelet plasma membrane exposing negatively charged phospholipids. The thrombin thus generated, further activates platelets and leads to fibrin deposition resulting in a stable impermeable haemostatic plug[6,7]. Platelet adhesion, activation and aggregation, together with concurrent thrombin generation are central events in this process.

1.7.5 VWF region D'D3 – FVIII complex formation

The FVIII binding site on VWF has been localised to the N-terminal 272 residues comprising domains TIL'E'VWD3 (aa 764-1035, V8 protease and tryptic digest fragment SPIII-T4)[54,237], of which the first 106 residues (aa 764-869) encompassing domains TIL'E' may be of particular importance[238]. The interaction of VWF with FVIII is of direct and great clinical significance, and the importance of VWF·FVIII complex formation is illustrated by patients with severe VWD (type 3), who have undetectable VWF levels. Not only do these patients have a concomitant deficiency in procoagulant FVIII, but they also have a considerably reduced half-life of intravenously administered FVIII[239,240]. This phenotype, that which mimics mild haemophilia A, is also observed in patients with type 2N VWD, harbouring *VWF* mutations that lead to decreased FVIII binding affinity[10,241]. This behaviour reflects the dependence of FVIII survival on the formation of VWF·FVIII complexes, and demonstrates the biological significance of this interaction.

Whereas VWF is synthesised throughout the vascular endothelium, the liver is likely to constitute the primary source of FVIII synthesis[242]. Synthesis has been localised to the cells lining the microvasculature of the liver, most likely sinusoidal endothelial cells and Kupffer cells[243]. It is of interest to note, that compared to FVIII, the liver is not a significant site of VWF synthesis, at least at the mRNA level[243]. Despite the association of VWF with FVIII in the peripheral circulation, there is no direct evidence to suggest that VWF and FVIII are synthesised together in any particular cell type *in vivo*[9]. Nevertheless, numerous indicators suggest that limited co-expression may exist. It is well recognised that the administration of DDAVP results in a transient increase in VWF and FVIII levels. Interestingly, recent studies have demonstrated that whilst liver transplantation cures haemophilia A, subsequent infusion of DDAVP in these patients produced a transient increase in plasma VWF levels, but did not further increase plasma FVIII levels[244]. Whereas non-haemophilic liver transplant recipients demonstrate responses in both VWF and FVIII following DDAVP. Furthermore, DDAVP administration does not significantly increase plasma FVIII levels in patients with type 3 VWD[10,245]. Cumulatively, these data support the hypothesis that a co-synthesised, releasable pool of VWF-FVIII may indeed exist *in vivo*[245].

FVIII circulates in blood as a heterodimer composed of two non-covalently associated polypeptide chains, a light chain with a molecular weight of 80 kDa and a heterogeneous heavy chain with a molecular weight varying between 90 and 210 kDa[239,246]. Immediately following release into the circulation, FVIII forms a tight non-covalent complex with VWF; each monomer in multimeric VWF is able to bind one FVIII molecule[247,248] with high affinity, the K_D of the interaction is in the sub-nanomolar range (0.2-0.5 nM)[247,249]. It is of note that high affinity binding is dependent upon the multimeric structure of VWF and assembly of heterodimeric FVIII, and it not known if all exposed FVIII binding sites on multimeric VWF must be occupied before binding to another VWF molecule. Kinetic analysis of VWF·FVIII complex formation reveals that it is characterised by very fast association and dissociation rate constants ($k_{on} \sim 7 \times 10^6$ mol^{-1} s^{-1} and $k_{off} \sim 4 \times 10^{-3}$ s^{-1})[250–252]. 50% of FVIII is bound to VWF in approximately two seconds and more than 95% will be bound within 12 seconds following FVIII cellular release[250,251]. This is extremely rapid in comparison to the half-life of a few hours observed for FVIII in the absence of VWF in plasma[240,250,253]. The relatively low stability of the VWF·FVIII complex, highlighted by the fast dissociation rate constant (Table 1.2), may be important for the regulation of FVIII activity; indeed, a high stability of the complex could hamper dissociation of FVIII from VWF following thrombin cleavage at sites of vascular injury[252]. An additional consequence of the high association and dissociation rates is that a small but significant proportion of FVIII molecules (5-8%) is unprotected by VWF[147,254].

The molar concentrations in plasma of the two molecules differ greatly. The typical plasma concentration of FVIII is 100-250 ng/mL (~1 nM, assuming an average molecular mass of 240 kDa), whereas that of VWF is ~10 μg/mL (~50 nM, in monomers). Therefore, *in vivo*, the stoichiometry is limited by the amount of FVIII, resulting in approximately a 1:50 ratio, such that not all VWF monomers are complexed with FVIII[247,250]. Even amongst normal individuals, plasma concentrations of VWF and FVIII vary with age, race, gender and ABO blood group. FVIII:Ag and VWF:Ag levels are significantly higher in African-Americans (approximately 20%) as compared with similar Caucasian populations, although the effect of ABO blood group is maintained. In addition, plasma FVIII and VWF levels rise with increasing age in

adults[255]. Importantly, elevated VWF and FVIII levels are independent risk factors for venous thromboembolism[204,205,256], thus the probability of deep vein thrombosis and pulmonary embolism is greater in non-O blood type patients and African-Americans[257].

FVIII is synthesised as a single polypeptide chain of 2351 aa that includes a 19 residue signal peptide and a 2332 residue mature protein (Figure 1.24). The FVIII protein also has a discrete domain organisation as follows: A1-a1-A2-a2-B-a3-A3-C1-C2[258,259]. The A domains are bordered by linkers (a1, a2 and a3) that contain clusters of Asp and Glu residues and hence are termed the acidic regions[239]. In the a3 region, residues 1649-1689 (41 amino acids), acidic residues comprise 37% of all amino acids. Within the Golgi apparatus, FVIII undergoes intracellular proteolysis. The middle and C-terminal regions of the B domain comprise a Arg-xx-Arg motif, that is similar to the motif recognised by intracellular proteases of the subtilisin-like family[246]. Consequently, the FVIII polypeptide chain is cleaved at Arg1313 and Arg1648 in the B domain, an event that disrupts the covalent linkage of the FVIII heavy chain (heavy chain, A1-a1-A2-a2-B) and light chain (light chain, a3-A3-C1-C2), giving rise to the heterodimeric molecule that circulates in plasma[239] (Figure 1.24). The heavy and light chains of FVIII remain non-covalently associated through the A1 and A3 domains in a divalent metal ion-dependant manner[260].

Mutations in the FVIII gene that cause aberrant biosynthesis or secretion result in haemophilia A, an X-linked bleeding disorder with an incidence of 1 in 5000 males. It is characterised by frequent and spontaneous bleeding into joints and soft tissues due to unstable clot formation[261]. In haemophiliacs, platelet surface FX activation by the FVIIIa·FIXa tenase complex (during propagation phase) is abolished and therefore, platelet surface thrombin generation fails. The TF·FVIIa complex cannot effectively substitute for FVIIIa·FIXa, as it produces FXa on TF-bearing cells, although it is possible for some FXa to diffuse through the fluid phase onto the platelet surface. In addition, platelet adhesion and aggregation at the site of injury occurs over the site of TF expression, resulting in an essentially insurmountable barrier to the movement of FXa (and other activated factors). Thus, fibrin production mediated by the extrinsic arm of coagulation alone is inadequate in both amount and distribution, to support haemostasis[6].

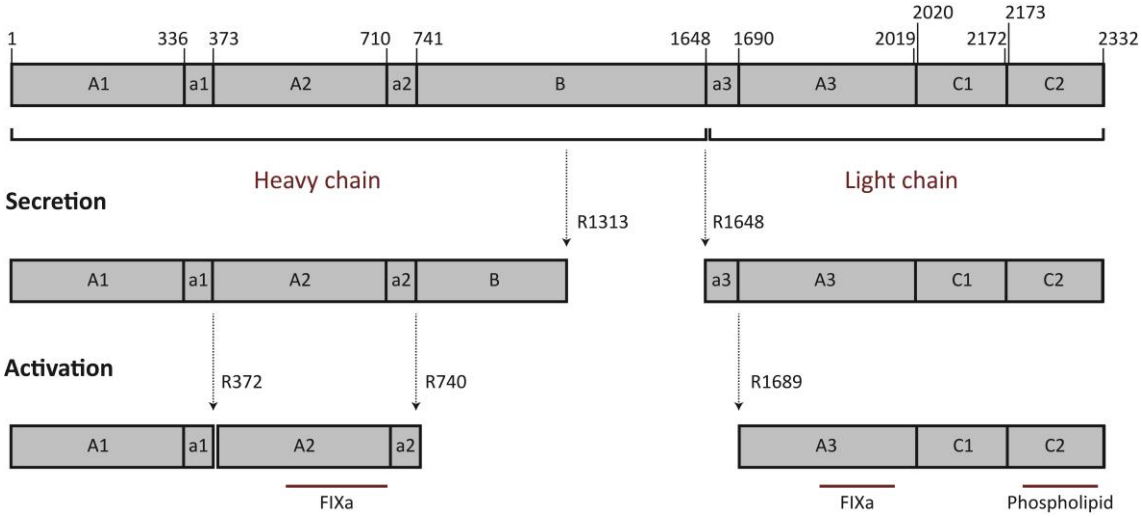


Figure 1.24: Factor VIII domain structure and arrangement of heavy and light chains. FVIII is a glycoprotein that acts as an initiator and regulator of the intrinsic pathway of coagulation. The N- to C linear arrangement of the heavy and light chains in shown in the top panel, the mature protein consists of 2332 residues which are arranged in a discrete domain structure. The A domains are bordered by acidic regions: a1 (aa 337-372), a2 (aa 711-740) and a3 (aa 1649-1689). FVIII is secreted as a heterodimer formed of a heavy and light chain (middle panel) due to intracellular proteolysis at Arg1313 and Arg1648, the latter event disrupts the covalent linkage between the heavy and light chains giving rise to heterodimeric FVIII. The heavy and light chains remain non-covalently associated through the A1 and A3 domains in a metal ion dependent manner. Immediately following release into circulation, FVIII interacts with its carrier protein VWF to form a tight, non-covalent complex. Two FVIII peptide regions are implicated in VWF binding, the light chain acidic a3 and C2 domains. Proteolytic activation of FVIII by thrombin (FIIa) proceeds by cleavage of the heavy chain at Arg372 and Arg740, and at Arg1689 in the light chain to yield the FVIIIa heterotrimer. FVIIIa dissociates from VWF, associates with the FIXa serine protease (via domains A2 and A3; FVIII light chain contributes most of the binding energy) and directs localisation of the resulting tenase complex to the membrane surface of activated platelets via an interaction with its C-terminal C2 domain. Membrane bound FVIIIa·FIXa activates FX, which then activates thrombin.

Although the assembly of the FVIII heterodimer increases the affinity of FVIII and VWF interaction, the VWF binding site on FVIII is entirely located within the light chain[248,249,262,263]. Two residue stretches on FVIII are implicated in binding VWF: the amino-terminal a3 acidic region of intact FVIII light chain and the C-terminal C1 and C2 domains. These regions are in close proximity and act synergistically in capturing VWF in plasma, with subsequent formation of the VWF·FVIII complex (K_D FVIII light chain binding to VWF, 3.8 nM)[247,264]. Independently, a3 and C2 present with affinities of 72 nM and 564 nM, respectively[264] (Table 1.2). However, it is the acidic a3 region that is critical to high affinity binding and this is dependent upon tyrosine sulphation at

residue 1680 of a3[265]. Mutation of this site (Tyr1680Cys/Phe) reduces the affinity for VWF by ~5-fold and is associated with mild to moderate haemophilia A[266]. In these patients, FVIII:C is fully functional but FVIII:Ag content is reduced to 10-20% of normal levels[267]. Furthermore, loss of acidic a3 from the FVIII heterodimer (heavy chain/A3-C1-C3) results in a >1000-fold decreased K_D for VWF (Table 1.2); this would predict a similar reduction for VWF affinity upon thrombin a activation[264].

Table 1.2: Kinetics of the interaction between VWF and FVIII [251,252,262,264,268–270]**.** LCh, light chain; HCh, heavy chain; SPIII-T4, VWF V8 protease and tryptic digest fragment.

Interaction	K_D	k_{on}	k_{off}
VWF·FVIII	0.2-0.5 nM	8.56×10^6 M^{-1} s^{-1}	6.89×10^{-3} s^{-1}
VWF·FVIII LCh	3.0-4.1 nM	9.2×10^5 M^{-1} s^{-1}	3.2×10^{-3} s^{-1}
VWF·A3C1C2	625 nM	6.4×10^4 M^{-1} s^{-1}	4.0×10^{-2} s^{-1}
VWF·C2	564 nM	6.2×10^4 M^{-1} s^{-1}	3.5×10^{-2} s^{-1}
VWF·a3 fragment	72 nM	3.3×10^4 M^{-1} s^{-1}	2.4×10^{-3} s^{-1}
VWF·HCh/A3C1C2	640 nM	5.93×10^4 M^{-1} s^{-1}	3.8×10^{-2} s^{-1}
Dimeric VWF·FVIII	4.95-6.37 nM	-	-
SPIII-T4·FVIII	48.5 nM	-	-
SPIII-T4·FVIII LCh	102 nM	0.81×10^5 M^{-1} s^{-1}	8.2×10^{-3} s^{-1}

Thrombin activates FVIII (FVIIIa) by cleavage at one site in the light chain, Arg1689, and at two sites in the heavy chain: Arg372 and Arg740. Cleavage at Arg1689 bisects the acidic a3 domain from the remainder of the light chain (Figure 1.24). Removal of domain a3 significantly reduces the affinity of FVIIIa for VWF, thereby releasing FVIIIa from the VWF·FVIII complex. Dissociation exposes FVIII surfaces that are important for efficient FVIIIa association with FIXa, and binding to phospholipid on the surface of activated platelets[264,271]. Mutation of Arg1689, as found in haemophilia A, predisposes to a haemorrhagic diathesis due to the inability of thrombin to cleave and dissociate FVIIIa, to assemble the FVIIIa·FIXa (tenase) complex[239]. Interestingly, in the crystal structures of B domain-less FVIII, the acidic a3 domain of the FVIII light chain could not be resolved and remained disordered, suggesting that the interaction with VWF may be important in stabilising this region[272,273].

In addition to the high affinity acidic a3 binding site on FVIII, the C2 domain, and to some extent the C1 domain, have also been shown to be important in determining interaction with the N-terminus of VWF. Interestingly, structural analysis of the FVIII light chain structure reveals that the C1 domain forms an extensive hydrophobic interface with domain A3 and is thus fixed, whereas domain C2 is only loosely tethered to the C1 domain and is capable of significant motion, thereby, possessing a greater degree of conformational flexibility[272,273]. Contributory regions to VWF binding within the FVIII C2 domain have been identified following characterisation of various recombinant C2 variants[274–276], including some haemophilia A associated mutations. In particular, the hydrophobic doublets Met2199/Phe2200 and Leu2251/Leu2252 have been shown to directly interact with VWF[275,277], as well as to phospholipid membrane, whereby cluster mutation of these residues resulted in a more than 50-fold reduction in VWF binding affinity[275] (Figure 1.25).

Figure 1.25: Hydrophobic spikes of FVIIII C2 domain. The VWF and membrane interactive hydrophobic amino acids in β-hairpin turns of FVIII light chain C2 domain are shown as red carbon-α spheres[275,277]. Residue numbers relate to the full length mature FVIII protein.

A role for the C1 domain in mediating the interaction between FVIII and VWF has also been described, a naturally occurring inhibitory antibody directed towards an epitope

within the C1 domain has been shown to inhibit VWF·FVIII complex formation and the mutations Ser2119Tyr and Arg2159Cys resulted in an 80- and 50-fold loss of affinity for VWF, respectively[278–281] (Figure 1.26). However, given the dramatic loss of affinity of the FVIII light chain for VWF following thrombin cleavage at Arg1689 and loss of the N-terminal acidic a3 sequence, the importance of the C1 and C2 domains seems somewhat incongruous. Nevertheless, specific residues within C1 and C2 may contribute to the avidity, rather than affinity, of VWF·FVIII complex formation.

Convergent lines of evidence suggest that VWF·FVIII complex formation is principally mediated by electrostatic forces[252]. First, the putative binding sites for VWF on the FVIII molecule are strongly and oppositely charged; the acidic a3 region is negatively charged, whereas the C2 domain bears an overall positive charge. Second, VWD and haemophilia A disease-causing mutations of charged residues in VWF and FVIII, respectively, have been shown to abolish binding[236,274,276,281,282]. Furthermore, perturbation of the electrostatic properties of charged C2 domain residues, underlies the inhibition mechanism of the VWF·FVIII interaction by pathogenic FVIII inhibitors[276]. Third, the presence of sulphated Tyr1680 on FVIII a3, which is able to establish electrostatic interactions, is indispensable for the binding of FVIII to VWF[265]. Lastly, it is well established that exposure to high salt concentrations leads to dissociation of the VWF·FVIII complex[252,283,284], and this is solely due to a decrease in the second-order association rate constant; the dissociation rate constant of the interaction is not significantly influenced by ionic strength[252]. Therefore, coupled with the high VWF·FVIII association rate constant at physiological ionic strengths (Table 1.2) that exceeds 10^5 M^{-1} s^{-1} (a value suggested as a start of the diffusion controlled regime of protein-protein interactions[285]), this implies that the formation of the VWF·FVIII complex depends on long-range electrostatic steering effects[252]. VWF·FVIII complex formation is also inhibited by acidic pH, with negligible interaction at pH values less than 5; this behaviour reflects the decisive role ionisation of charged residues in FVIII play in VWF interaction and attest to a predominately electrostatic interaction[252].

Type 2N VWD represents a relatively rare form of VWD, estimated to affect ~1-5 people per million population, and accounting for approximately 1-2% of all identified VWD cases (13% of all type 2 VWD)[286,287]. Recessive inheritance of two mutations is

necessary for type 2N VWD. Affected individuals are either homozygous for a single type 2N mutations, compound heterozygous for two different 2N missense mutations or compound heterozygous for a type 2N mutation and a silent VWF allele. The N-terminal 272 residues of VWF, containing the FVIII binding site, is encoded by exons 18-23; however, 87% of all type 2N missense mutations reported occur in exons 18-20 encoding domains TIL'E' (D' region)[10,236]. Type 2N mutations usually occur within hypermutable (CpG dinucleotide sequence) arginine codons, and the missense mutation Arg854Gln (c.2561G>A) occurs with polymorphic frequency in Caucasian populations[10,288].

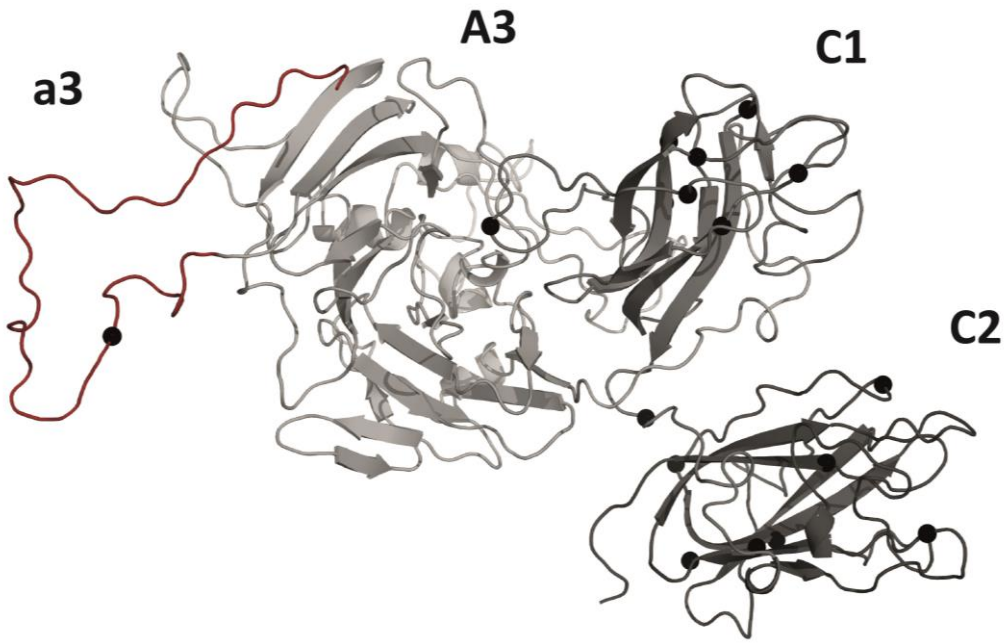


Figure 1.26: Haemophilia A mutations in the FVIII light chain that decrease VWF binding. The acidic a3 domain (red) is an all loop region and this disorder prevented its assignment in the reported FVIII crystal structures. Thus the full length FVIII light chain was modelled as a chimera using crystal structures of FVIII (pdb: 3CDZ)[272] to model domains A3-C1-C2 and an *Escherichia coli* kinase (pdb: 1OJ4) for the a3 region, in Modeller9v8. A ribbon representation of the full length FVIII light chain is shown, with the a3 region coloured red and the A3-C1-C2 domains in progressively darker shades of grey. Reported mutations[9,274,276,281] that have been shown to decrease VWF binding are represented by carbon-α spheres in black. a3: Y1680C/F. C1: Q2087E, R2090C, I2098S, S2119Y, N2129S, R2150C/H, P2153Q, and R2159C. C2: S2173I, A2201P, V2232A, P2300L, R2304C/G/H/L, R2307Q. Interestingly, two mutations in domain C2 show enhanced VWF binding: D2288A and R2320T.

Consequent to the intrinsic defect in VWF that results in impaired FVIII binding and complex formation, plasma FVIII is labile, prone to early proteolysis and levels are usually lower than those of plasma VWF[287]. It is well established that interaction of

FVIII with VWF increases survival of FVIII in plasma. The half-life of infused FVIII concentrate in patients with type 3 VWD is only 1-2.5 hours, as compared to ~12 hours in patients with haemophilia A[9,240,253]. Thus, VWF·FVIII complex formation is critical to the survival of FVIII in circulation; free FVIII is rapidly cleared from the circulation whereas, complexed FVIII decays with the half-life of VWF[289]. The mechanism for maintaining FVIII circulatory residence time includes stabilisation of the FVIII structure and the prevention of cleavage, and removal by cellular interactions. When FVIII is bound to VWF, a stable association of its heavy chain and light chain is maintained[290]; interaction of the FVIII light chain with VWF enhances the rate of light and heavy chain association and may function as a surface upon which the FVIII heterodimer assembles[262,291]. VWF protects FVIII from proteolytic degradation by phospholipid-binding serine proteases such as activated protein C (a potent anticoagulant) and activated FXa[289,292–294]. Furthermore, VWF bound FVIII cannot bind to FIXa, thereby denying FVIII the capacity to assemble the FX-activating tenase (FVIIIa·FIXa) complex[295–297]. The protection of FVIII by VWF is mediated by two mechanisms. First, VWF-bound FVIII is competitively inhibited from binding to phospholipids or platelets[269,298], second, direct protease-binding sites within the FVIII light chain are cryptic whilst FVIII is in complex with VWF[297,299]. Furthermore, rapid VWF·FVIII complex formation, may prevent FVIII from interacting with components that bind with lower affinity[250], such as FIXa (K_D 15 nM)[295] and phospholipids (K_D 2 to 4 nM)[300]. This protection from proteolysis serves to increase the FVIII circulatory residence time.

As well as concealing binding sites on FVIII essential to its procoagulant activity, VWF interaction modulates clearance of FVIII from the circulation by macrophage low-density lipoprotein receptor-related protein-1 (LRP1). VWF inhibits uptake of FVIII by macrophage LRP1 by shielding the LRP1 binding site on the FVIII light chain[301,302]. VWF does not interact with LRP1 in the absence of shear[301–303], and due to the higher affinity of FVIII for VWF (K_D FVIII interaction with LRP1 ~60 nM)[301], VWF prevents binding of complexed FVIII to the macrophage receptor[9]. Intriguingly however, VWF does not protect FVIII from LRP1-dependent endocytosis under conditions of high shear, which converts sheared VWF into its active platelet-binding conformation and a

ligand for LRP1, and thus, shear flow drives the catabolism of the VWF·FVIII complex by macrophages[303,304].

Moreover, complex formation with VWF has a direct effect on the immunogenicity of FVIII, and may protect FVIII against inhibitor formation in haemophilia patients. Inhibitor formation against infused FVIII is among the most serious complications of haemophilia treatment, occurring in ~25% of patients[305]. It is a T-cell-dependent immune response to foreign infused FVIII, rendering it ineffective and leads to poorly controlled bleeding and severe morbidity. The light chain C2 domain is a significant antigenic target of anti-FVIII antibodies and a frequent epitope has been localised to residues 2248 through 2312[306]. Antibodies to this epitope have been shown to inhibit FVIII binding to both VWF and phospholipid, but are less reactive to FVIII in complex with VWF. In addition to masking epitopes on FVIII, complex formation also prevents FVIII interaction with antigen presenting cells such as macrophages and dendritic cells and thereby blocking subsequent activation of $CD4^+$ T-cells[9,305].

The molecular structure of the FVIII binding region on VWF is not known, and the importance of its elucidation is two-fold. First, it would provide important and detailed insights into the mechanism of FVIII docking onto VWF, and the minimal unit required to stabilise FVIII *in vivo*. This would allow for the improvement of haemophilia A treatment that relies on FVIII replacement therapy, as well as the development of novel approaches to treat both haemophilia A and VWD. Plasma-derived and recombinant human FVIII products are used for the treatment of haemophilia A. Following intravenous infusion, the primary determinant of FVIII residence time in plasma is its non-covalent association with VWF. However, haemostatic protection with current FVIII products is temporally limited as a consequence of a short FVIII half-life of ~8-12 hours, requiring frequent prophylactic injection to maintain FVIII at a level that is protective against spontaneous bleeding episodes (above 1%)[307,308]. Thus, targeting FVIII half-life extension would have the greatest impact in achieving long therapeutic windows with protective plasma levels in haemophilia A patients. Second, analysis of VWD patient mutations against a backdrop of both structural and functional perturbations will provide the link between genetic pathology and clinical phenotype, thereby enabling improved and targeted management of VWD patients.

1.8 NMR for protein structure determination

NMR spectroscopy, first described by Felix Bloch and Edward Mills Purcell in 1956, is the phenomenon of atoms absorbing radio frequency electromagnetic radiation (RF) under the influence of a magnetic field. For NMR spectroscopy of biomolecules, the detection of this absorption requires the use of isotopes of hydrogen, nitrogen and carbon that possess odd mass numbers and a spin ½.

1.8.1 Chemical shift

Spin is a quantum mechanical property attributed to protons, neutrons and electrons, measured in multiples of ½ and can be positive or negative. The most important nuclei with a ground state (I) spin ½ are ^{1}H, ^{13}C and ^{15}N. A nuclear state with spin I is $(2I + 1)$-fold degenerate, thus a nucleus with spin ½ will have two possible states - a low or high energy state. In the absence of a magnetic field the two states are of equal energy, however, in an applied magnetic field the energies of the spin states are split into two nuclear spin levels: nuclear Zeeman splitting (Figure 1.27).

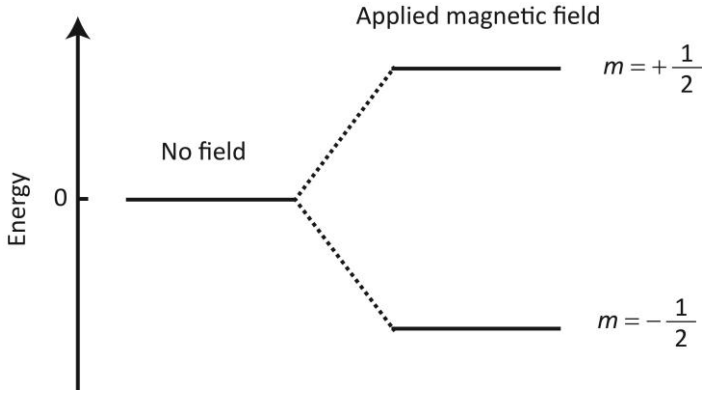


Figure 1.27: Energy levels for a nucleus with spin ½. The nuclear Zeeman sublevels of a spin ½ nucleus are shown. Each level is given a magnetic quantum number, m.

At equilibrium, transitions from one spin state to another occur infrequently. However, the application of RF energy equal to the difference between the two nuclear states (the Larmor frequency) will promote the oscillation between the two states. Following a pulse of RF energy, the spin states of the majority nuclei will relax to their ground state. This relaxation is known as free induction decay (FID). The energy required to flip nuclear states is dependent upon the strength of the magnetic field, the type of

nucleus and the surrounding chemical environment, and this gives rise to a range of resonance frequencies for each nucleus, an effect termed the chemical shift. The chemical shift mechanism is due to the applied magnetic field inducing currents in the electron clouds of the molecule, these circulating molecular currents, in turn, generate an induced magnetic field. The nuclear spins experience the total effect of the applied magnetic field as well as the induced field generated by molecular electrons, and hence give rise to measurable shifts in nuclear resonance frequencies.

Chemical shift is an extremely important concept in NMR as it permits the identification of individual nuclei and their subsequent characterisation. Because chemical shift reflects the local electronic environment of the nucleus, it contains information about protein structure, wherein secondary structure elements are associated with specific chemical environments of the spin. As a result, the secondary structure of folded proteins can be identified from their chemical shifts. A near complete chemical shift assignment is a prerequisite for structural studies by NMR.

1.8.2 J-couplings

When nuclei are bonded to one another, the magnetisation may be transferred through the bonding electrons. This type of spin-spin coupling is termed scalar or J-coupling and it provides a direct spectral manifestation of the chemical bond; unlike the chemical shift, the J-coupling independent of the applied magnetic field. Two spins have a measurable J-coupling only if they are linked together through a small number of chemical bonds, including hydrogen bonds. The J-coupling is exclusively intramolecular. Protein structure determination relies on J-couplings to identify the sets of resonances of individual amino acid residues in the NMR spectra and to subsequently establish the sequential connectivity of amino acids in the protein sequence.

1.8.3 Nuclear Overhauser enhancement

Whereas J-coupling is a through-bond transfer of magnetisation, nuclear Overhauser enhancement (NOE) is a measurement of through-space transfer. The NOE is a manifestation of the prediction that dipolar-coupled spins do not relax independently and is one of the most important effects in NMR spectroscopy. As each nuclear spin is a magnetic dipole, it generates a magnetic field around itself (Figure 1.28, spin I). A second nuclear spin (spin S) in spatial proximity may interact with the magnetic field of spin I. The corollary is that spin I will also experience the effect of the magnetic field generated by spin S. Such an interaction is a direct dipole-dipole coupling and thus leads to the NOE.

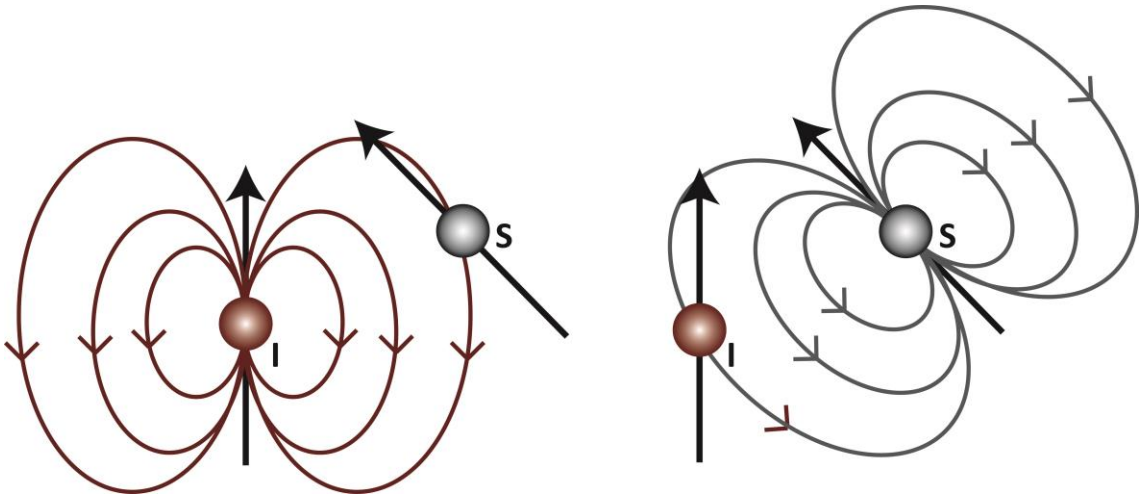


Figure 1.28: Magnetic fields generated by nuclear spins I and S. This through space magnetisation transfer allows distance information to be obtained for both intra- and inter-residue interactions between protons that are typically less than 6 Å apart.

The NOE is currently the most significant NMR parameter in protein structure determination, as it provides a measurement of through space interactions between atoms that are spatially close ($\leq$ 6 Å). An NOE occurs when magnetisation is transferred between two protons through saturation of absorption intensity of the spins followed by dipolar cross relaxation. The dipolar cross relaxation rate constant is proportional to the inverse sixth power of the distance between the two interacting protons as given by the following equation:

$$\sigma_{IS} = r_{IS}^{-6}\gamma_I\gamma_S f(\tau_C)$$

Where σ_{IS} is the rate constant of magnetisation transfer, r_{IS} is the distance between the two interacting nuclei, γ_I and γ_S are gyromagnetic ratios of the two nuclei and τ_C is the correlation time for molecular tumbling. A majority of the restraints used to calculate the three dimensional structure of a protein come from many hundreds of NOE cross-peaks, and are represented as bounds on the separation of pairs of ^{1}H atoms. Thus, NOEs are used to restrain the molecular structure and conformation unambiguously.

1.8.4 Residual dipolar couplings

The r^{-6} dependence of NOEs stipulates that the distance-dependent information provided is both short-range and local; the presence of two pairs of NOEs does not provide insight into how they are related to one another. In contrast, the effect of dipole-dipole coupling may be used to obtain distance-independent orientational restraints, termed residual dipolar couplings (RDCs). RDCs provide additional structural information regarding the orientation of inter-atomic vectors with respect to a molecular coordinate frame.

In isotropic solutions the dipolar coupling between two spins averages to zero due to molecular motion; however, to extract dipolar coupling data the molecule must behave anisotropically. When molecular tumbling is anisotropic, such as when a protein is partially aligned in a liquid crystal, the orientational distribution is not equiprobable and the molecule has a preferential orientation, thus the dipolar coupling does not average to zero. This non-zero coupling is called the RDC.

There are several types of alignment media used to generate partial alignment of biological molecules for RDC measurements. They include bicelles, filamentous bacteriophage *Pf*1 and poly(ethyleneglycol)/hexanol mixtures. These co-solutes, restrict free molecular tumbling and induce alignment by steric (bicelles and poly(ethyleneglycol)/hexanol mixtures) or electrostatic (*Pf*1) interactions.

Although partial structure determination can be performed with RDC data alone, they are most often used as restraints to refine existing structures, as well as to determine the relative orientation of domains in multi-domain proteins.

1.9 Summary

VWF is a Janus-faced protein that is critical to the processes of both haemostasis and its pathologic extensions bleeding and thrombosis (Figure 1.29). The cysteine-rich N- and C-terminal regions of VWF have been the subject of significant efforts aimed at characterising the structure and plasticity of function of this unique protein, particularly in view of the involvement of VWF in pathological processes that transcend haemostasis. The primary aim of this work is to determine a high-resolution structure of the FVIII binding region on VWF that would reveal insights into the mechanism of VWF·FVIII complex formation and the molecular basis of the bleeding diathesis type 2N VWD.

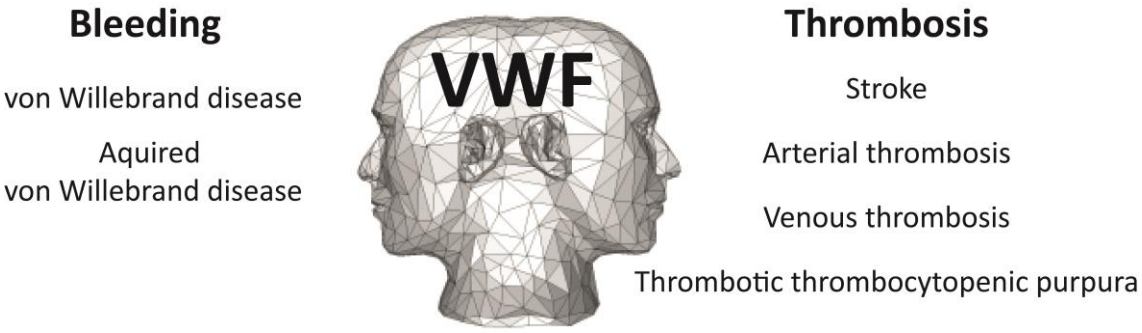


Figure 1.29: VWF: a Janus protein in bleeding and thrombosis. The haemostatic system is designed to react rapidly upon vascular damage to minimise blood loss, a process that requires tight regulation to prevent thrombotic occlusion of the vessels. VWF is a key player in the haemostatic system, as illustrated by its liaison with both haemorrhagic and thrombotic conditions.

2. Results and Discussion

Electron microscopy studies have given a first and fascinating glimpse into the domain arrangement and overall structure of the multi-domain VWF protein. The limited resolution of this technique, however, has precluded the elucidation of vital atomic details needed to examine the molecular recognition and binding of FVIII to VWF. Therefore, detailed analysis of the VWF molecular structure that facilitates binding to FVIII, which would reveal crucial mechanistic insight into their intimately intertwined activities, has until now, remained elusive.

Solving the three-dimensional structure of the biological macromolecule under investigation is essential for the complete understanding of its function and of the structure-function correlation. Unlike X-ray crystallography, NMR spectroscopy permits the determination of not only an atomic resolution molecular structure in solution, but also of time-dependent phenomena such as protein dynamics. Function in biological systems is exquisitely dependent upon spatial and temporal changes in the protein structure, and these dynamics contribute to the stability of functional states and play an important role in catalysis, where conformational rearrangements can juxtapose key residues; in ligand binding, which may require structural rearrangements; in molecular recognition processes, which are often fine-tuned by disorder-to-order transitions; and in allostery, where coupled structural fluctuations can transmit information between distant sites in a protein[309]. Hence, NMR is uniquely suited to the study of both protein structure and site-specific dynamic processes.

The FVIII binding region on VWF is contained within the amino-terminal 272 residues (exons 20-23)[54,237] comprising domains, TIL' (aa 767-827), E' (aa 829-863) and VWD3 (aa 867-1031). Several lines of evidence, however, suggest that of these regions, domains TIL' and E' (the classical D' region) are essential for FVIII binding. First, 72% of unique missense mutations that directly affect FVIII binding to VWF (type 2N VWD) are found in domains TIL'E'. The majority of the remaining type 2N mutations are in domains VWD3 and C83. Importantly, type 2N mutations that result in the most severe phenotypes and the lowest circulating FVIII levels, are found in the TIL' domain[236,310–312]. Second, the epitopes of the most potent FVIII binding-blocking monoclonal

antibodies (mAb) map to the TIL' domain; the epitopes of mAb 418 and mAb 32B12 encompass residues 765-816 and 814-823 in TIL', respectively[313–315]. Third, the interaction between VWF and the FVIII light chain is disrupted in high ionic strength solutions and by low pH (< pH 5), indicating that VWF·FVIII complex formation is mediated by electrostatic interactions[252,283]. Domains TIL'E' contain numerous charged residues that likely contribute to this interaction; there is an over-representation of charge-change inducing type 2N mutations in TIL'E'.

Despite the importance of the interaction of VWF with FVIII under physiological and pathological conditions, the molecular mechanisms that govern VWF·FVIII complex formation remain incompletely defined. The regions of VWF and FVIII involved in binding, as well as the affinity and kinetics of the interaction have been investigated. However, the absence of structural information has precluded the understanding of the detailed topology and mechanism of VWF·FVIII complex formation. The results discussed herein, are of the successful expression of the VWF TIL'E' (D') region in *Escherichia coli* (*E. coli*) and the determination of its high resolution NMR structure and dynamic characteristics. The TIL' domain lacks extensive secondary structure, is strikingly dynamic and is coincident with the clustering of pathological mutations leading to decreased FVIII binding affinity (type 2N VWD). The major FVIII binding site is localised to a specific and flexible pocket in domain TIL', that is supported by the structured regions of TIL' and the rigid E' domain.

2.1. Determination of TIL'E' domain boundaries by sequence analysis

The application of NMR spectroscopy to structure determination requires the correct assignment of VWF domain boundaries to ensure that the structure of an autonomously folded TIL'E' region is solved. The TIL' and E' domains were delineated by positionally conserved TIL (*Ascaris*) and fibronectin I-like domain family cysteines. The TIL'E' fragment spans VWF residues 767 through 863 (97 aa), encompassing sixteen cysteines that form eight disulphide bonds (Figure 2.1). Three of five cysteine linkages in domain TIL' have been biochemically characterised[42]; the remainder, together with the three disulphide bonds in E', were assigned by the positional

conservation of cysteine residues in a multiple sequence alignment of homologous domain sequences.

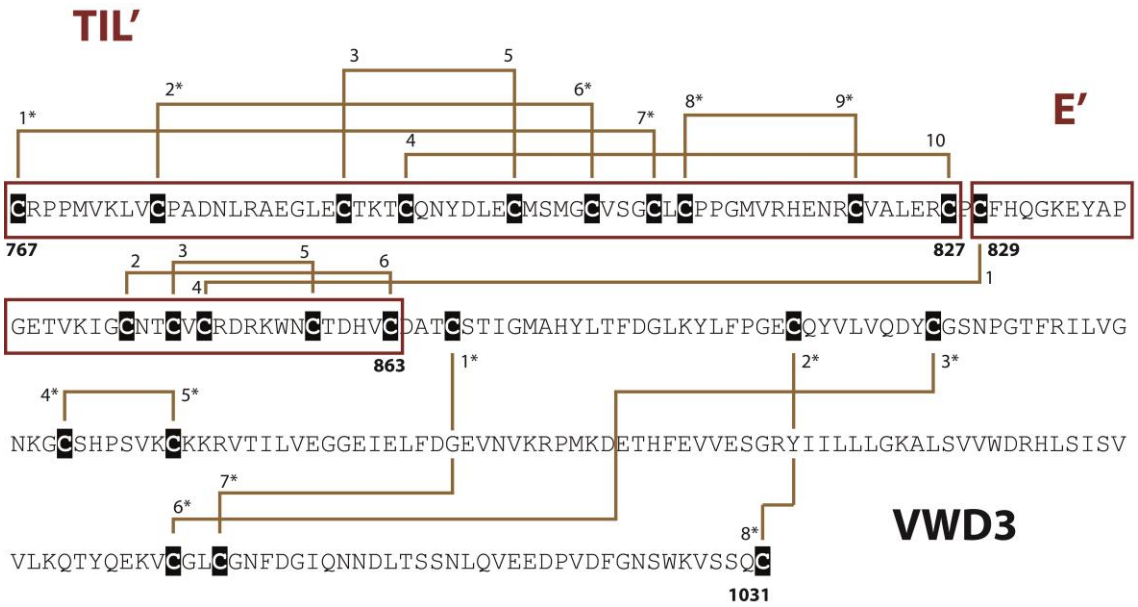


Figure 2.1: FVIII binding sequence on VWF. The FVIII binding region on VWF, identified within a tryptic fragment termed SPIII-T4 (aa 764-1035)[54], is formed of three conserved domains, TIL', E' and VWD3, the sequence of which is shown. Conserved Cys residues are highlighted in black and their connectivity is shown by gold bars. Biochemically determined disulphide bridges[42] are indicated by a * and these are, TIL': Cys767-Cys808 (1-7), Cys776-Cys804 (2-6) and Cys810-Cys821 (8-9); VWD3: Cys867-Cys996 (1-7), Cys889-Cys1031 (2-8), Cys898-Cys993 (3-6) and Cys914-Cys921 (4-5).

The disulphide bond topology is well preserved in the *Ascaris* family, their Cys residue connectivity is as follows: Cys_1-Cys_7, Cys_4-Cys_{10}, Cys_3-Cys_5, Cys_2-Cys_6, Cys_8-Cys_9 (Figure 2.2, panel A); and correspond to VWF TIL' disulphide bridges Cys767-Cys808 (1-7), Cys776-Cys804 (2-6), Cys788-Cys799 (3-5), Cys792-Cys827 (4-10) and Cys810-Cys821 (8-9). Disulphide bond linkages of the TIL domains, determined initially by chemical methods alone (1-9, 3-7, 2-5, 4-6, 8-10)[110], were revealed to support neither the structure determined by crystallography[105,316], nor assignments derived from NMR spectra[104,317–319], highlighting the difficulty in unambiguously pairing such closely spaced Cys residues. A multiple sequence alignment of the conserved ten-cysteine core of the TIL domain family (~60 aa) was generated, which enhanced the reliability of domain boundary assignments, and facilitated the alignment of VWF TIL' with its structural homologues for subsequent comparative modelling (Figure 2.2, panel A).

The *Apis mellifera* chymotrypsin/cathepsin G inhibitor-1 NMR structure[317] (AMCI-1; pdb: 1CCV) displays significant sequence identity with VWF TIL' (32.3% identity and 37.1% similarity); modelling based on this template required the least number of insertions and deletions in a pairwise alignment (Figure 2.2, panel B). Regions of sequence insertion or deletion corresponded to accessible loop regions of structure 1CCV, as is desired for optimal modelling.

The E' domain was assigned based on sequence conservation of the VWC subdomain 1 cysteine signature Cys_1-Cys_2xxCys_3xCys_4-Cys_5-Cys_6 (Figure 2.2, panel C). The fibronectin I-like family comprises small ~35 residue domains, and scanning by means of fold recognition algorithms[320], against all structures deposited in the protein data bank, ascertains their structural homology to the N-terminal VWC domain fold[118]. In the ten-cysteine full VWC domains, conserved residues Cys_2 and Cys_6 participate in 2-8 and 6-9 linkages, respectively, in which the Cys_2-Cys_8 linkage couples subdomain 1 to subdomain 2. Free thiol groups are rare and the reported pK_a values of Cys thiols are in the region of 8-9, thus their low abundance is likely to be linked with the risk of aberrant polymerisation an ionised cysteine may present at extracellular basic pH values[321,322]. As such, homology modelling of E' was based on the reasonable hypothesis that Cys_2 and Cys_6 would not remain unbound but rather, form a Cys_2-Cys_6 disulphide bridge in agreement with the close proximity of the corresponding residues in fibronectin I domains[323], resulting in the following linkage pattern: Cys829-Cys851 (1-4), Cys846-Cys863 (2-6) and Cys849-Cys858 (3-5) (Figure 2.2 panel C). The VWF E' domain was modelled against the NMR structure of VWC subdomain 1 of human collagen IIA (pdb: 1U5M)[118], sharing 30.6% and 38.9% sequence identity and similarity, respectively (Figure 2.2, panel D). Therefore, based on sequence homology, these comparative models provide a reliable prediction of the overall structure of the VWF TIL' and E' modules, and of the complete TIL'E' disulphide bond topology (Figure 2.2).

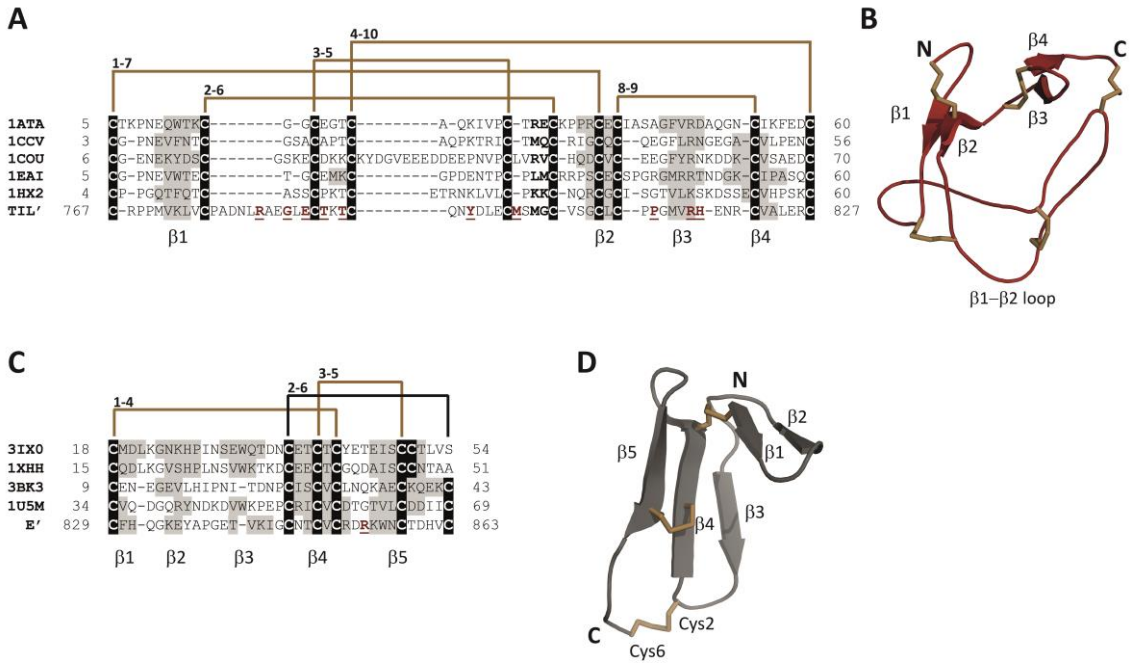


Figure 2.2: Sequence and structural modules of TIL'E'. Sequence alignments of VWF TIL' and E' with their structural homologues are shown (pdb codes are given). Conserved cysteines are highlighted in black, regions of β-strands are shown in grey. β-strand positions are based on the high resolution structures of TIL and fibronectin I-like family domains, whereas those of VWF TIL' and E' are derived from secondary structure assignment (DSSP) of the homology models. Positions of non-cysteine type 2N von Willebrand disease mutations are shown in red and underlined in the sequence. Panel A shows an alignment of the *Ascaris* protease inhibitor family whose disulphide bond connectivity is shown in gold and residues that would constitute the reactive site P1 and P1' scissile bond in this family are in bold. In panel B is shown the TIL' homology model based on structure 1CCV[317]. Panel C shows an alignment of members of the fibronectin 1-like family and modelling of domain E' on subdomain 1 of VWC structure 1U5M[118] is shown in panel D - the induced Cys2-Cys6 disulphide bond is indicated.

The correct formation of disulphide bonds is of critical importance to the maintenance of structure and function, as well as to protein stability by reducing the entropy of the unfolded state[324,325]. It has long been noted that disulphide bonds are highly conserved in evolution[322,326], and interestingly, in almost all cases conservation of half-cysteines even surpasses that of tryptophan, which is usually the most conserved amino acid[326]. It is striking that instances where the bond is not conserved, pairs of half-cysteines are replaced in concert in 99% of cases[327]; such is the case for the Cys_3-Cys_5 bond in VWF C domains VWC2, VWC4 and VWC6 (Introduction, Figure 1.8). Similarly, a novel single VWC domain family has recently been characterised in arthropods[328] that is also distinguished by its selective loss of disulphide bridge 3-5, which is arguably the least

important bond as it connects two β-strands within a sheet. Importantly, the Cys_2-Cys_6 (Cys846-Cys863) linkage predicted in domain E' is thus, corroborated by its specific absence in domain E2 (Introduction, Figure 1.6). The disulphide bonding pattern of both the TIL' and E' domains cannot be separated it into two discrete subpatterns of independently cross-linked regions (Figure 2.3, panels A and B). This connectivity formation is far more frequent than an independent disulphide bonding topology[329] (Figure 2.3, panel C), and it is thought to confer evolutionary advantage, as entropic stabilisation is greater for inseparable patterns whose complexity may be essential for maintaining the protein fold[330]; in support of the conservation of TIL' and E' disulphide bond topology.

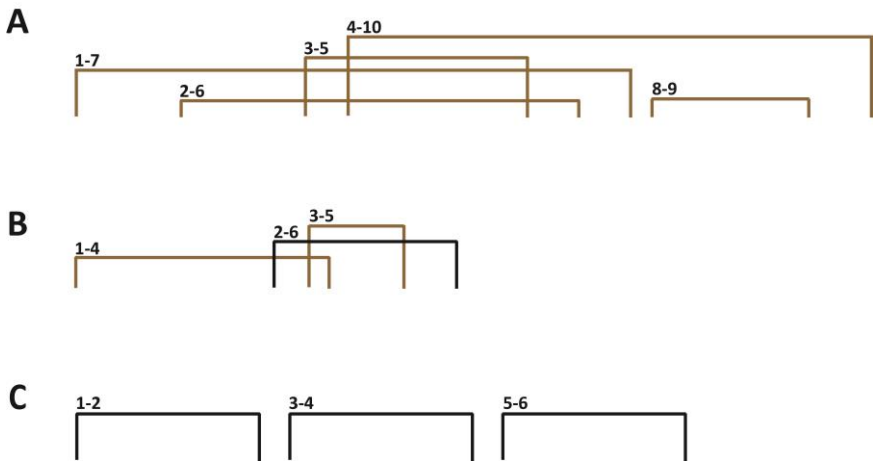


Figure 2.3: Disulphide bonding patterns of domains TIL'E'. Connectivities are composed of disulphide cross-linked loops. In the relative arrangement of disulphide bonds, there are three relationships that may occur between any two cross-linked loops; overlap, enclosure and independence[330]. Panel A shows the disulphide connectivity of domain TIL', in which there is overlap and disulphide bridge 8-9 is enclosed within the loop of disulphide bridge 4-10. Panel B shows the overlapping disulphide bonding pattern of domain E'. Panel C shows an example of an independent disulphide bonding pattern which can be divided into discrete cross-linked regions.

Similar disulphide connectivity patterns place similar spatial geometric constraints on the protein backbone, and consequently on the overall three-dimensional structure[327,331,332], and this facilitated prediction of the discrete TIL' and E' domain structures by virtue of homology (Figure 2.2 panels B and D). Protein structures diverge linearly, although at a much lower rate, with the sequence variability in homologous proteins. Hence, structure is more conserved than sequence[333,334]. However, when the protein core is stabilised by disulphide bonds, as is the case for small domains (≤ 70

residues) such as TIL' (61 residues) and E' (35 residues), disulphide bond conservation with sequence divergence can be reasoned to selection pressure preserving geometrical restraints[333,335]. Even in the case of extremely low sequence similarity (sequence identify < 25%)[331], there exists an intimate relationship between disulphide connectivity patterns and protein structure. Based on this premise, together with the knowledge that disulphide bonds very rarely occur between protein domains[322], the reliability of using cysteine signatures to delineate VWF domain boundaries and build comparative models is greatly increased. The TIL' and E' domain boundaries form the basis for biochemical and structural characterisation of the FVIII binding region.

2.2. Rational design of TIL'E' expression constructs

A preliminary assessment of the TIL' and E' multiple sequence alignments and homology structures guided the choice of N- and C-termini end residues chosen for expression. This prevented both the inadvertent truncation of conserved regions and disruption of predicted secondary structural elements in TIL'E'. Additionally, this approach limited the addition of excess flexible N- and C-termini loop residues that give rise to poorly resolved cross peaks in NMR spectra and do not contribute structural information. Initially, numerous expression constructs and expression conditions were tested. Ultimately, the thioredoxin (TRX)-tagged VWF TIL'E' fragment, comprising residues 766-864 was chosen for expression and subsequent structural assignment by NMR.

The production of correctly folded protein in sufficient quantity and purity is essential for advancement to structural studies; the enteric bacterium *E. coli* remains the most popular host for heterologous expression of recombinant proteins. However, the cytoplasm of wild type *E. coli*, is a reducing environment and is therefore, unable to sustain disulphide bond formation, which requires the pairwise oxidation of free thiol groups present on Cys residues; indeed cysteine is the most abundant amino acid in proVWF, comprising 8.3% of all residues and 16% of the TIL'E' domains. Given 2*n* cysteines that form *n* disulphide bonds, the number of possible connectivity patterns of all 2*n* cysteines is: $(2n)! \,/\, (2^n \cdot n!)$, thus as the number of cysteines increases, the

number of potential pairings increases factorially[327,336]. For the 16 cysteines of TIL'E', there are 2,027,025 possible combinatorial disulphide patterns. To circumvent this problem, an *E. coli* mutant strain with a cytoplasmic redox potential comparable to that of the mammalian endoplasmic reticulum has been developed by disruption of the thioredoxin and glutathione reductase genes. This *trxB*$^{-}$/*gor*$^{-}$ double mutant is conducive to the formation of disulphide bonds in the cytoplasm[337]. Nevertheless, protein overexpression remains particularly arduous when disulphide bonding is both abundant and topologically complex, such is the case for TIL'E', and this poses a significant challenge to recombinant production by *E. coli*.

When proteins with more than two cysteines are oxidised, non-native disulphide bonds inevitably form, which results in a randomly scrambled set of protein isomers. Proteins that are mis-oxidised and therefore mis-folded may be isomerised back to their native disulphide-bonded state during *in vivo* folding by the *E. coli* periplasmic oxidoreductase chaperone DsbC. The isomerase activity of DsbC is of critical importance to the non-consecutive, multi-disulphide bonded VWF TIL'E' fragment, hence, the oxidative *E. coli* strain SHuffle[337], engineered to co-expresses active cytoplasmic DsbC was used for the expression of thioredoxin tagged TIL'E' (VWF residues 766-864).

Fusion of TIL'E' to an N-terminal thioredoxin tag was essential for successful expression of the protein in *E. coli*. Notably, the thioredoxin tag resembles the native VWF prodomains that contain thioredoxin-like sequences[93]. Although, TIL'E' (residues 766-864) in the absence of an N-terminal thioredoxin tag (Figure 2.4, panel A) could be expressed and purified from *E. coli*, incorrect disulphide bonds were formed. The purified TIL'E' domain (lacking a thioredoxin tag) migrated as a doublet when analysed by non-reducing SDS-PAGE (Figure 2.4, panel B lanes 6-10 and panel C). As this protein migrated as a single band under reducing conditions (Figure 2.4, panel B lanes 1-2 and panel C), this indicates that multiple disulphide-bonded species (isomers) were present.

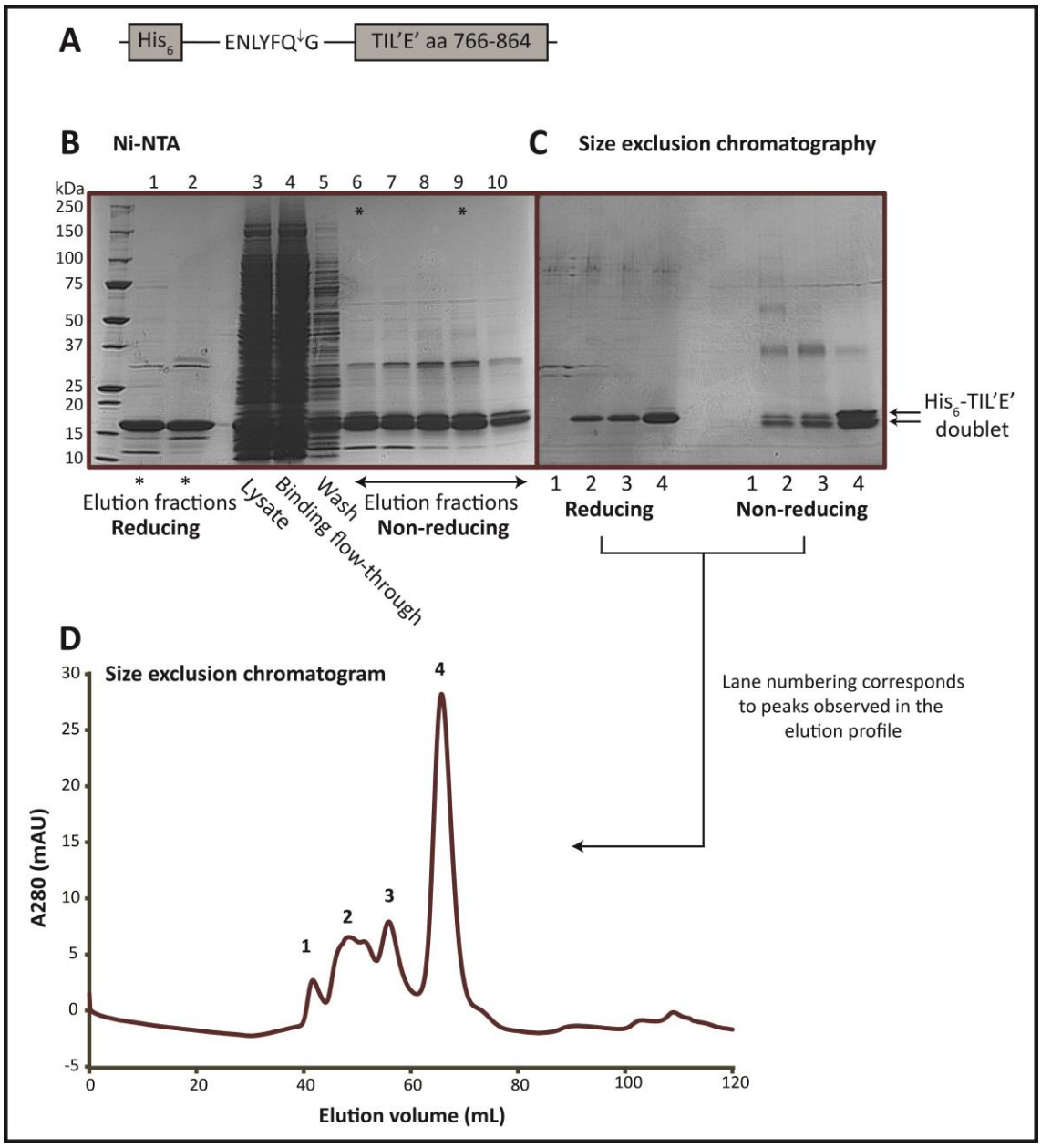


Figure 2.4: Purification of His_6-TIL'E'. A schematic of the His_6-TIL'E' construct is shown in panel A. The linker between the His_6 tag and TIL'E' encompasses the TEV cleavage site (ENLYFQ$^{\downarrow}$G). This purification is of uncleaved His_6-TIL'E'. Panel B shows the SDS-PAGE of fractions taken from the initial Ni-NTA (Nickel, Ni^{2+}) affinity column purification. The lysate, binding flow-through and washing fractions are run under non-reducing conditions and are shown in lanes 3-5. Ni-NTA elution fractions (lanes 6-10) show the bound His_6-TIL'E' protein (13.2 kDa) that migrates as a doublet under non-reducing conditions. His_6-TIL'E' elution fractions in lane 6 and 9 (*), are also run under reducing conditions in lanes 1 and 2, in which the doublet collapses and migrates as a single band, indicating the presence of disulphide isomers. Fractions containing His_6-TIL'E' (lanes 6-10) are then applied over a size exclusion chromatography column (16/60 Superdex 75), the elution fractions of which are shown in panel C under reducing and non-reducing conditions. The lane numbering corresponds to peaks 1 to 4 of the chromatogram shown in panel D. Under non-reducing conditions, purified His_6-TIL'E' migrates as a doublet.

Details of the final VWF TIL'E' construct are given in Materials and Methods (section 4.3) in brief however, the TIL'E' coding sequence (Ser766-Asp864, 99 residues) was cloned into a pET32b(+) vector containing an N-terminal thioredoxin and His_6 tag. The TRX-His_6 tag is separated from the TIL'E' coding sequence by a 31 residue linker containing the Tobacco Etch Virus (TEV) cleavage sequence (ENLYFQ↓G) to allow removal of the tag from purified TIL'E' (Figure 2.5).

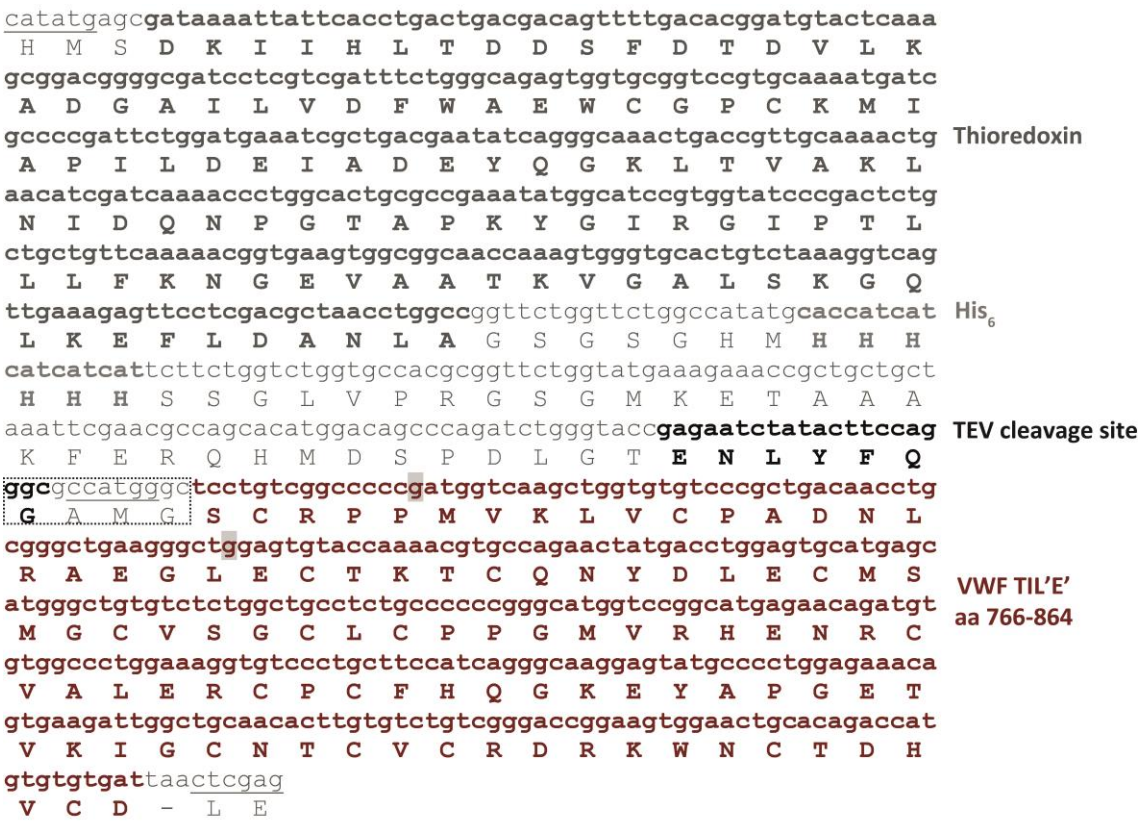


Figure 2.5: Arrangement of construct TRX-His_6-TIL'E'. The nucleotide sequence of the TRX-His_6-TIL'E' construct and its amino acid translation is shown. The 107 residue thioredoxin tag (dark grey, bold) is followed by a six histidine tag (light grey, bold) and the TEV protease cleavage site (ENLYFQ↓G, in black bold). VWF residues Ser766 through Asp864 (cDNA bp 2296-2592) were cloned using restriction sites 5' *NcoI* (ccatgg) and 3' *XhoI* (ctcgag), these are underlined in the sequence. Cleavage by TEV leaves four residues (GAMG, boxed) N-terminal to VWF Ser766 as cloning artefacts. Highlighted in grey are single base pair mutation sites in TIL' to remove an *NcoI* (c.2310C>G) and an *XhoI* (c.2358C>G) restriction site within TIL'. The stop codon TAA follows VWF residue Asp864.

Once suitable domain boundaries that define a folded TIL'E' region were established and the expression construct and conditions optimised, uniformly isotopically labelled TIL'E' was successfully obtained by over-expression of TIL'E' in M9 minimal medium containing ^{15}N-ammonium chloride ($^{15}NH_4Cl$) and $^{13}C_6$-D-glucose. TIL'E' was subsequently purified to homogeneity by a two-step Ni-NTA (nickel) affinity

chromatography purification protocol. TRX-His_6-TIL'E' is eluted from the first Ni-NTA column application following binding of the His_6 tag (Figure 2.6, panel A lanes 3-6), the TRX-His_6 tag is subsequently cleaved by TEV and the cleavage reaction (Figure 2.6, panel B lane 1) reapplied to the Ni-NTA column, from which cleaved TIL'E' is collected in the flow-through. Soluble monomeric TIL'E' is then applied to a size exclusion chromatographic column as a final purification step, from which pure TIL'E' reproducibly elutes at 73 mL, in a single symmetrical peak (Figure 2.6 panel C). Its subsequent migration as a single band, under non-reducing SDS-PAGE, indicates that the recombinantly produced TIL'E' is homogenous and monomeric (Figure 2.6, panel B lane 7). The calculated polypeptide masses for tagged (TRX-His_6-TIL'E') and untagged TIL'E' are 28.6 kDa and 11.3 kDa, respectively. The final TIL'E' protein contains 103 residues of which the N-terminal four amino acids (GAMG) constitute cloning artefacts (Figure 2.5).

Of interest, mutation of the N-linked glycosylation site Asn857, in domain TIL' and adjacent to Cys858, results in increased levels of free thiol groups in VWF when expressed in mammalian cell lines and hence, is not compatible with formation of the correct VWF disulphide bond architecture[145]. However, unglycosylated TIL'E' expressed in an oxidising cytoplasm in *E. coli*, was able to form the correct disulphide bonded structure. This may be reasoned to the continual exposure of TIL'E' to an oxidative DsbC-enhanced folding environment, that does not rely on an N-glycan dependent calnexin/calreticulin folding cycle[338] or glycan chain mediated interaction with endoplasmic reticulum protein disulphide isomerase ERp57, which has been shown to be involved in VWF synthesis[145,339].

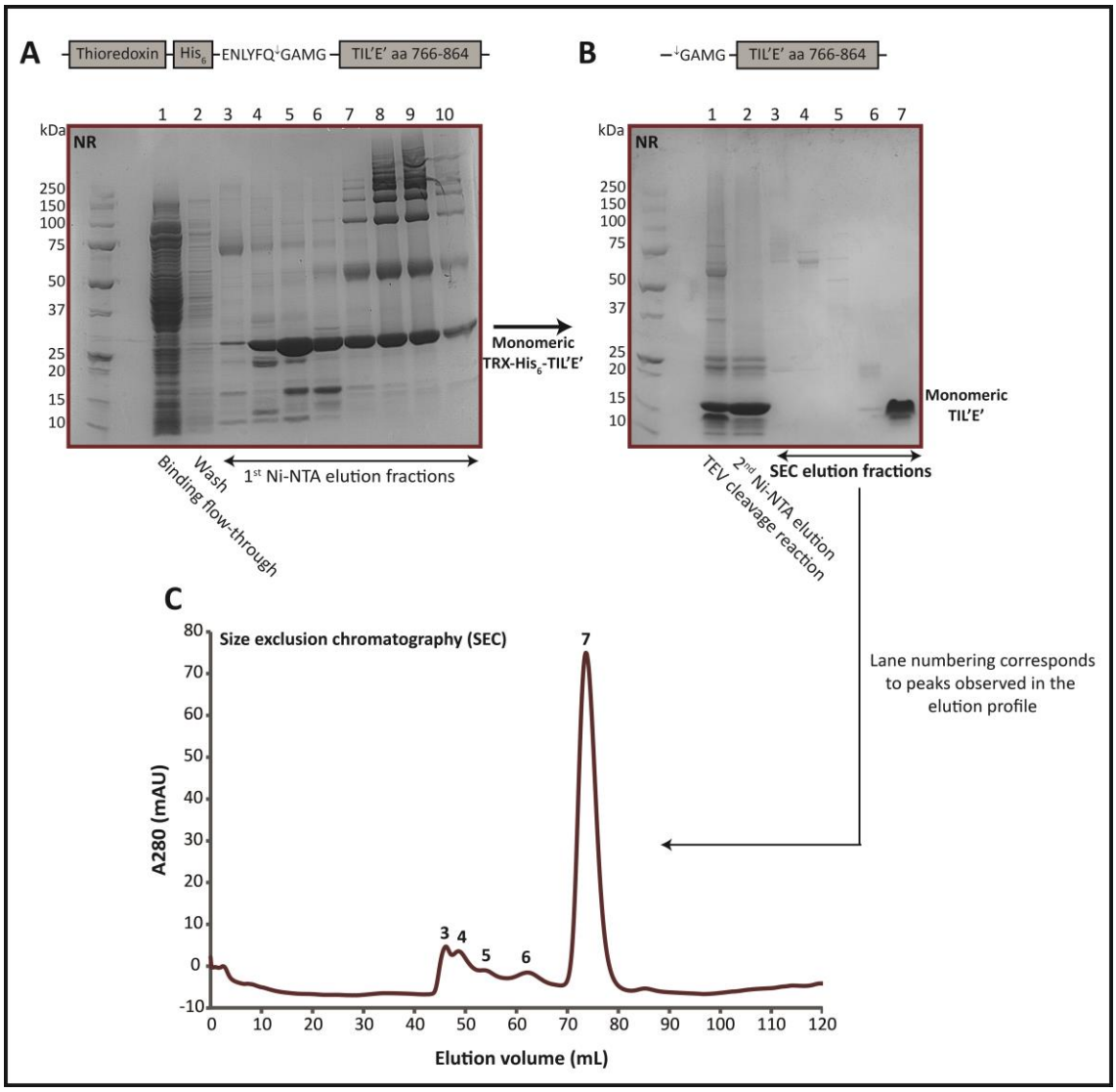


Figure 2.6: Purification of TRX-His$_6$-TIL'E'. Shown in panel A is a non-reducing (NR) SDS-PAGE of the first Ni-NTA affinity purification step. Above the SDS-PAGE is a schematic of the construct at this stage of the purification. Lane 1 shows the unbound material from the filtered *E. coli* cell cytoplasm (binding flow-through) and lane 2 shows the washing flow-through. Lanes 3 to 10 are successive elution fractions of bound TRX-His$_6$TIL'E'. Monomeric TRX-His$_6$-TIL'E' migrates as a single well-defined band and elutes as expected for a ~29 kDa protein. It is clearly seen that fractions containing higher molecular weight TRX-His$_6$-TIL'E' forms (lanes 7-10) with incorrectly formed disulphide bonds are segregated from predominantly monomeric TRX-His$_6$-TIL'E' fractions. Monomeric TRX-His$_6$-TIL'E' fractions (lanes 3-6) are collected and subjected to Tobacco Etch Virus (TEV) protease for proteolytic cleavage of the TRX-His$_6$ tag, the reaction of which is shown in lane 1 of panel B, where TEV appears as a doublet of ~21 kDa and the cleaved TRX-His$_6$ tag obscures the TIL'E' domain (11.3 kDa). A schematic of the cleaved construct is shown above the SDS-PAGE in panel B. Cleaved TIL'E' does not bind to the second Ni-NTA column as it does not contain a His$_6$ tag and is retrieved in the flow through, concentrated and subsequently applied to a 16/60 Superdex 75 size exclusion column. The peak fractions eluting from the size exclusion chromatography (SEC) step are shown in panel B lanes 3-7 and are numbered in the order in which they elute from the column, as shown in the chromatogram in panel C. Pure, monomeric TIL'E' (11.3 kDa) consistently elutes at 73 mL in a single symmetrical peak (peak 7) and migrates as a single band under non-reducing SDS-PAGE as shown in panel B lane 7.

2.3. TIL'E' expressed in *E. coli* is correctly folded and disulphide bonded

Following purification of monomeric TIL'E', Ellman's assay and nanoelectrospray ionisation mass spectroscopy (nanoESI-MS) were used to show that TIL'E' is correctly folded and contains no free thiol groups. A compact folded protein is expected to have a significantly smaller projected area compared to unstructured protein molecules, and would therefore accommodate fewer charges on the protein surface during the ion evaporation stage. This phenomenon is reflected in the native nanoESI-MS profile of TIL'E' (Figure 2.7), in which compelling evidence that the TIL'E' protein is well-folded is clearly provided by the very narrow charge-state distribution, centred on the $[M+6H]^{6+}$ state.

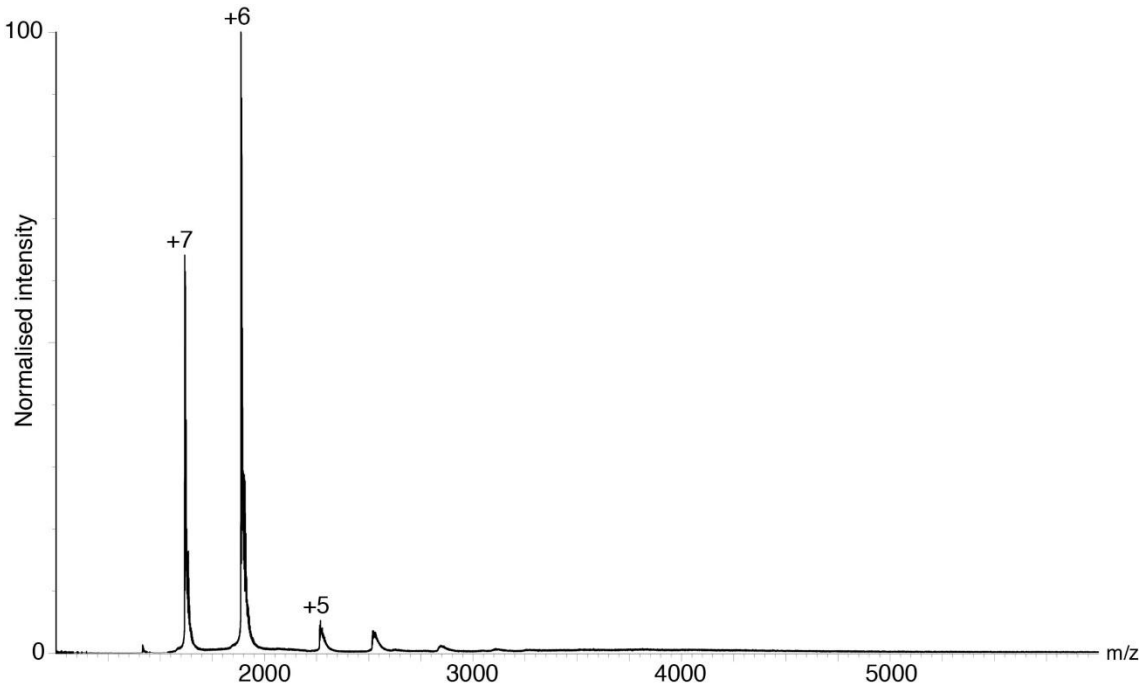


Figure 2.7: Native nanoESI-MS of recombinantly produced $^{1}H^{12}C^{14}N$ TIL'E'. The ESI-MS profile of TIL'E' shows a very narrow charge-state distribution as expected for a well-folded protein molecule. An unfolded protein would display a far broader charge-state distribution than that shown above, due to the accommodation of more positive charges during the ion evaporation stage. The experimental mass obtained from this MS is 11318.90 $\pm$ 0.12 Da. The calculated mass with eight disulphide bonds formed is 11320.19 Da, while with no disulphide bonds the calculated mass would be 11336.32 Da. The ESI-MS profile reveals a significantly narrow projected area, as expected for a well-folded protein molecule. Mass spectrometry and generation of this figure were kindly undertaken by Dr Konstantinos Thalassinos.

In addition, a 1D ^{1}H NMR experiment was used to further evaluate the conformational properties of TIL'E'; the folded or unfolded (unstructured) state of a protein may be rapidly deduced from a 1D ^{1}H NMR spectrum. Within the 1D profile, very sharp methyl peaks coupled with a lack of dispersion of the backbone amides (H^N) clustered in the region of 7-8.5 ppm, suggesting that all H^N protons are in a similar (averaged) chemical environment, is strongly indicative of an unstructured protein. The 1D spectrum of TIL'E', however, exhibits a broad $^{1}H^N$ chemical shift dispersion, especially downfield of 8.5 ppm, in addition to isolated methyl group signals below 0.5 ppm, together these are sensitive and reliable probes for a folded protein structure (Figure 2.8).

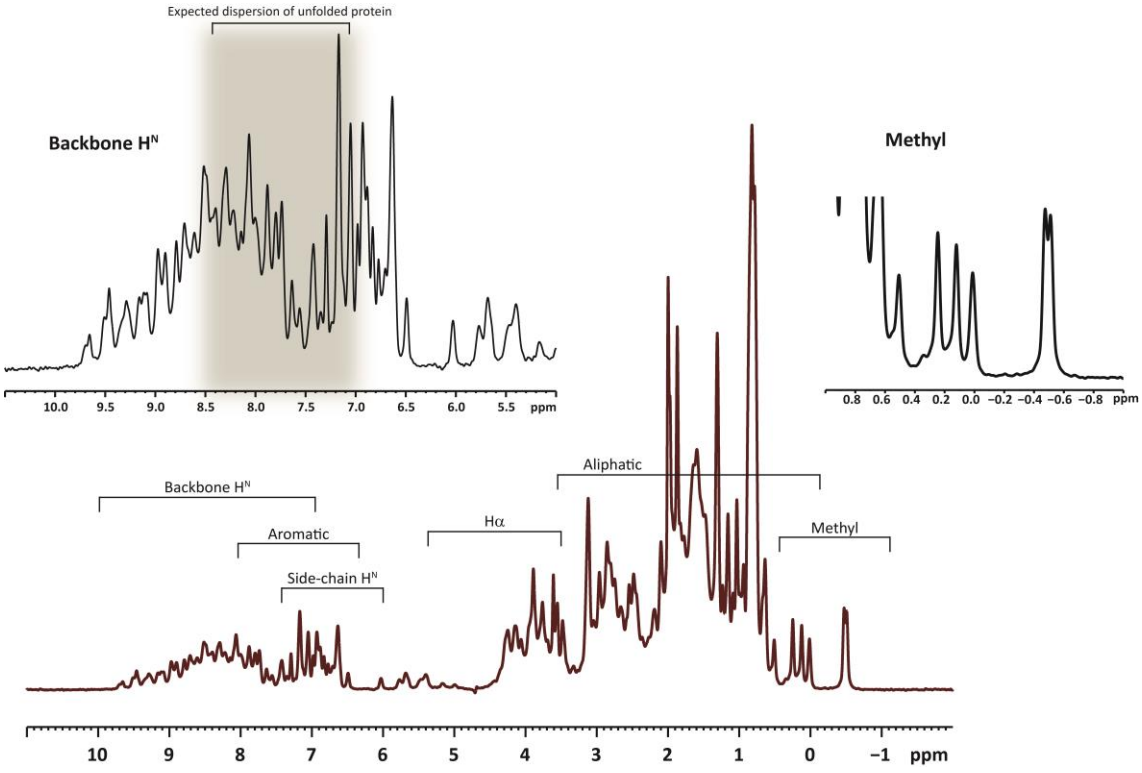


Figure 2.8: 1D ^{1}H NMR spectrum of TIL'E'. The spectrum was recorded in 20 mM sodium phosphate, 100 mM NaCl, pH 7.4, 25°C and at 500 MHz. The x-axis represents the chemical shift, a frequency scale expressed in parts per million (ppm). Shown are the chemical shift ranges observed for various types of ^{1}H resonances. An enhanced view of the backbone amide region is shown in the upper left and a broad chemical shift dispersion is seen, an area highlighted in grey between 7 ppm and 8.5 ppm indicates the region in which backbone amide peaks would cluster if TIL'E' had been unstructured. The upper left corner shows the methyl group region, highlighting the distinct methyl peaks below 0.5 ppm indicative of a folded protein.

To evaluate the conformation of TIL'E' at a residue level, a 2D ^{15}N-HSQC (heteronuclear single quantum coherence) spectrum was recorded of ^{15}N-labelled TIL'E'. An HSQC

experiment correlates each nitrogen atom of an amide to the attached proton, and thus the spectrum contains a cross peak for each backbone amide group (^{1}H-^{15}N), except proline, for a well behaved protein. Also visible in the HSQC spectrum are the NH_2 side chain groups of Asn (Nδ_2-Hδ_2), Gln (Nε_2-Hε_2) and the Trp indole side chains (Nε-Hε). In essence, a ^{15}N-HSQC spectrum provides an excellent high-resolution fingerprint of TIL'E' (Figure 2.9).

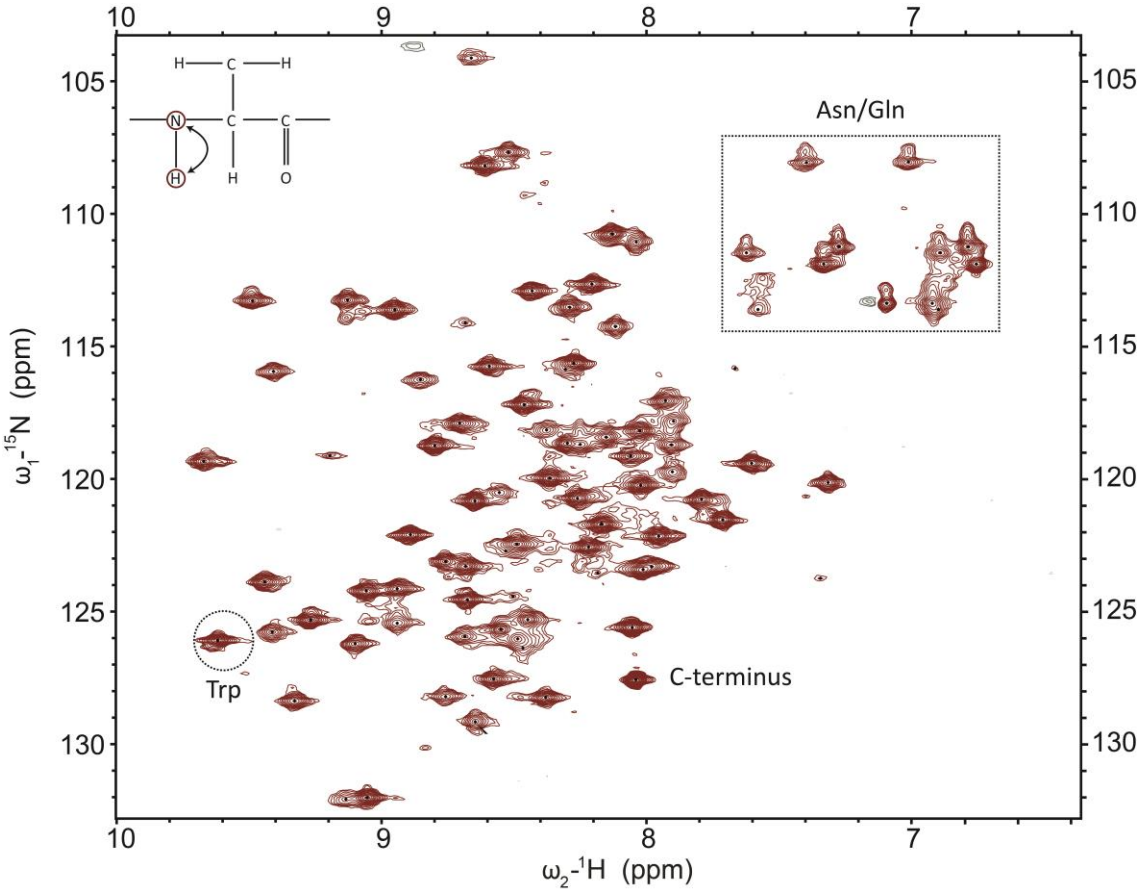


Figure 2.9: 2D ^{15}N-HSQC NMR spectrum of ^{13}C^{15}N TIL'E'. The TIL'E' ^{15}N-HSQC spectrum, showing all ^{1}H-^{15}N (amide group) correlations, was recorded at 600 MHz in 20 mM sodium phosphate, 100 mM NaCl, pH 7.4, 25°C. A schematic of the magnetisation transfer is shown in the top right corner of the spectrum. Most of the peaks are well dispersed, indicating the TIL'E' protein is well folded. The expected single tryptophan side chain Nε-Hε group is circled and the asparagine/glutamine side chain Nδ-Hδ_2/Nε-Hε_2 groups are boxed in the top right corner. The C-terminus of the protein usually gives an intense peak in the bottom left of the spectrum. 79 of the expected 93 peaks (excluding seven prolines and including the indole of Trp856) are seen. The absence of a peak may be due spectral overlap, loop dynamics or conformational exchange.

In corroboration of the 1D TIL'E' spectrum (Figure 2.8), the many discrete and dispersed peaks of the ^{15}N-HSQC spectrum are indicative of a stable tertiary structure (Figure 2.9). Of the 99 residues in TIL'E', seven are proline, therefore, including a

resonance for the indole ^{1}H-^{15}N of Trp856, 93 ^{1}H-^{15}N peaks are expected in the ^{15}N-HSQC spectrum; however, only 79 peaks are observed. Considered collectively, with the inhomogeneous peak intensities observed at inspection of the HSQC spectrum, the decreased number of ^{1}H-^{15}N peaks are an initial indication that the folded TIL'E' domains contain flexible and dynamic regions. Furthermore, when all cysteine residues are reduced, a starkly different HSQC spectrum is obtained, in which there was complete unfolding of TIL'E' as demonstrated by the significantly reduced ^{1}H-^{15}N peak dispersion, ranging between 8 and 8.5 ppm (Figure 2.10, black spectrum); thereby, highlighting the folded configuration of the native TIL'E' molecule (Figure 2.10, red spectrum) and the structural importance of the TIL'E' disulphide bonds in maintaining the fold.

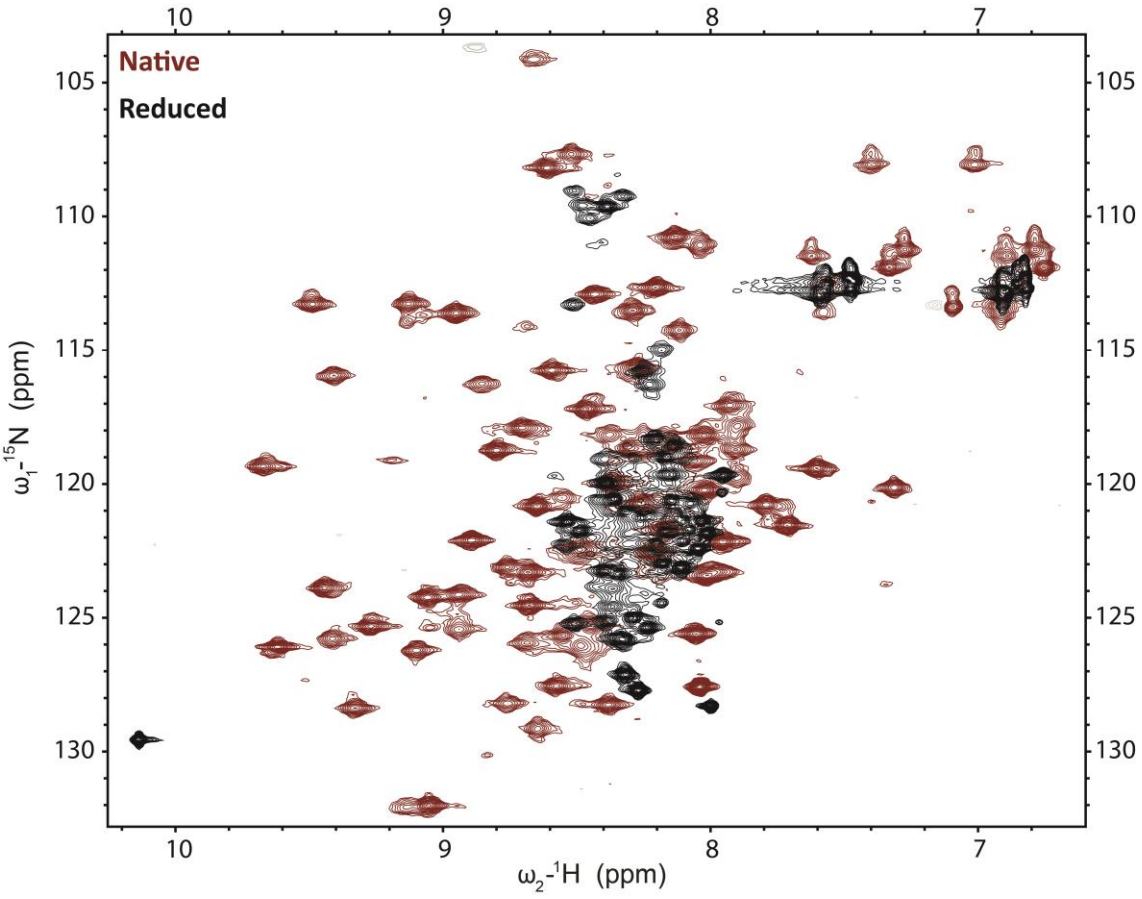


Figure 2.10: 2D ^{15}N-HSQC spectrum of reduced ^{15}N TIL'E'. In red is shown the native spectrum of TIL'E' over which lays the HSQC spectrum of reduced TIL'E' (black) in 20 mM sodium phosphate, 100 mM NaCl, 10 mM DTT, pH 7.4. The spectra were recorded at 600 MHz and 25°C. Reduction of TIL'E' with 10 mM DTT results in the collapse of the amide signals into the centre of the spectrum between 8 and 8.5 ppm. A classic spectrum of an unfolded protein.

It is noteworthy, that in contrast to the TIL'E' molecule (VWF residues 766-864), expression and purification of thioredoxin-tagged TIL' domain alone (residues 766-828) revealed that in the absence of E', it is unstable and cannot achieve a natively folded configuration; forming an intermediate that is reminiscent of a molten-globule. In the HSQC spectrum of ^{15}N-labelled TIL', the non-native, but not unfolded state is evidenced by the extremely broad and overlapped peaks with an agglomeration of signals in a narrow range between 7.8-8.8 ppm (Figure 2.11). This suggests that there is substantial conformational exchange and complex dynamic behaviour. A number of sharper, dispersed peaks are present at chemical shift values above 8.6 ppm, indicating that TIL' is, to a certain extent, folded (Figure 2.11); unlike the unfolded HSQC spectrum of reduced TIL'E' in which peak dispersion in the ^{1}H dimension are not present (Figure 2.10). The dispersed TIL' peaks do not overlay with any peaks of the native folded TIL'E' HSQC spectrum (Figure 2.9) and therefore, it is not known which regions of TIL' are likely to be folded. The inability of TIL' to achieve a stable, native configuration suggests that the final TIL' structure cannot be assembled in the absence of domain E', such that there may be cooperativity in folding between the domains.

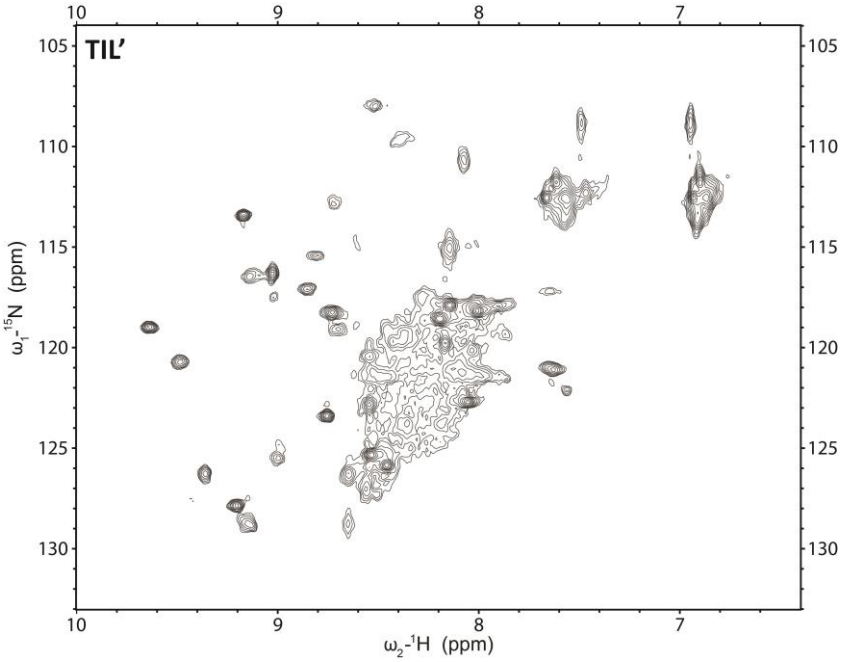


Figure 2.11: 2D ^{15}N-HSQC spectrum of ^{15}N TIL'. The TIL' ^{15}N-HSQC spectrum was recorded at 600 MHz in 20 mM sodium phosphate, 100 mM NaCl, pH 7.4, 25°C. TRX-His_6-TIL' was successfully expressed and domain TIL' (residues 766-828) purified following cleavage of TRX-His_6 tags. The fingerprint resonances from the amide protons in domain TIL' are not well dispersed, range predominantly from 7.8-8.8 ppm and display significant broadening and overlap. There are a few very sharp dispersed peaks, indicating that TIL' is in a non-native conformation.

In summary, recombinantly expressed TIL'E' is homogeneous, monomeric and contains no free thiol groups. It is of the expected molecular mass for an eight-disulphide bonded molecule and forms a stable tertiary structure. Furthermore, the formation of disulphide isomers when TIL'E' is expressed in the absence of a thioredoxin tag, together with the instability of domain TIL' when expressed alone, emphasise the importance of the continuity of the N-terminal proVWF domains in the correct folding and stability of TIL'E'.

2.4. TIL'E' binds to FVIII and mAb 418

The paragon of TIL'E' function is binding to FVIII. This is dependent upon the folded TIL'E' adopting its biologically active conformation. Proteolytic digestion of full length VWF with *staphylococcus aureus* V8 protease and trypsin yields a monomeric, N-terminal 272 residue fragment, termed SPIII-T4 (VWF residues 764-1035)[54,237]. The FVIII binding site on VWF has been shown to be contained within fragment SPIII-T4, comprising domains TIL'E'VWD3. The epitopes of extensively characterised anti-VWF antibodies, which block FVIII binding to VWF, have been mapped to residues Leu765-Arg816[237,313] (mAb 418, TIL'), Met814-Ala823[314] (mAb 32B12, TIL') and Cys829-Gly839[314] (mAb 31H3, E'), as shown in Table 2.1.

Table 2.1: FVIII binding-blocking anti-VWF antibodies

Monoclonal antibody	Epitope	Domain	VWF recognition
mAb 418	765-816	TIL'	Potent inhibitor of VWF·FVIII interaction, conformation-dependent (does not bind reduced VWF)[237,313]
mAb 32B12	814-823	TIL'	Potent inhibitor of VWF·FVIII interaction and recognises reduced VWF (conformation-independent)[314]
mAb 522[313]	798-844	TIL'-E'	-
mAb 31H3	829-839	E'	Moderate (partial) inhibitor of VWF·FVIII interaction, conformation-dependent, only recognises non-reduced VWF[314].
MAb W5-6A	841-859	E'	Moderate (partial) inhibitor[340]

The most potent inhibitors of the VWF·FVIII interaction, however, localise to domain TIL' (mAb 418 and mAb 32B12), with partial inhibition of FVIII binding achieved by mAb 31H3 binding to domain E'. This suggests that VWF domain TIL' forms the principal FVIII binding face.

Dot blot analyses of TIL'E' with recombinant FVIII were undertaken to validate retention of biological activity. Recombinant FVIII was spotted onto a nitrocellulose membrane in increasing amounts, from 125 ng to 1 μg, and incubated with TRX-His_6-TIL'E' (0.79 mg/mL); binding of TIL'E' to immobilised FVIII was subsequently detected with anti-His antibodies (Figure 2.12).

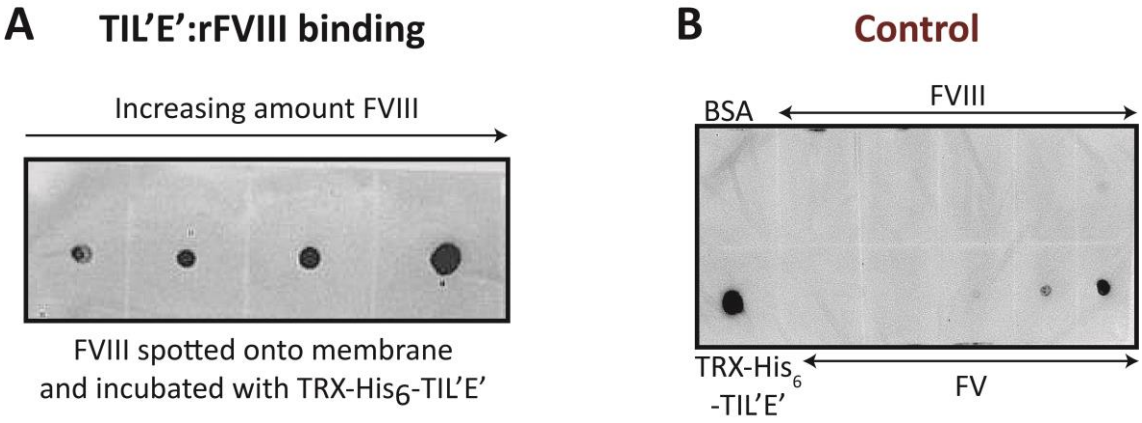


Figure 2.12: Dot blot analysis of TIL'E' functionality – binding to FVIII. The experimental blot is shown in panel A. Reconstituted, recombinant FVIII (Helixate NextGen 0.25 μg/IU) was spotted onto a nitrocellulose membrane in increasing amounts of 125 ng, 250 ng, 500 ng and 1 μg. The membrane was subsequently incubated with blocking buffer containing 0.79 mg/mL TRX-His_6-TIL'E'. Binding of TRX-His_6-TIL'E' to immobilised FVIII was revealed by ECL chemiluminescence detection of horse radish peroxidase (HRP)-conjugated anti-(penta)His antibodies. It is clear that TIL'E' binds to FVIII. A control blot was generated, shown in panel B, which confirms that there is no erroneous binding of anti-His_5 antibodies to immobilised FVIII. Spotted onto the control blot are a negative BSA control (250 ng), FVIII (62.5 ng to 1 μg) and FV (29 ng to 230 ng), a structural homologue of FVIII but does not bind to VWF. Surprisingly, FV binds directly to anti-His antibody at 57.5 ng, 115 ng and 230 ng, although the FV amino acid sequence does not contain three or more consecutive His residues. A positive TRX-His_6-TIL'E' control is shown in the bottom left corner of the control blot.

Binding of TIL'E' to the potent FVIII binding-blocking mAb 418[313] was also evaluated by dot blotting. The epitope of mouse anti-VWF mAb 418 (residues 765-816) encompasses the greater majority of domain TIL' and recognition of TIL' is sensitive to the reduction of VWF cysteines; mAb 418 is thus, a native VWF conformation-dependent antibody[237,313,314,341]. TIL'E' (0.45 mg/mL) was reduced with dithiothreitol

and alkylated with an excess of iodoacetamide. Both native TIL'E', and reduced and alkylated TIL'E' were blotted onto a nitrocellulose membrane in increasing amounts from 56.25 ng to 450 ng. The membrane was subsequently incubated with mAb 418 and binding of mAb 418 to TIL'E' was detected with a secondary anti-mouse antibody (Figure 2.13); mAb 418 binds only to native TIL'E'.

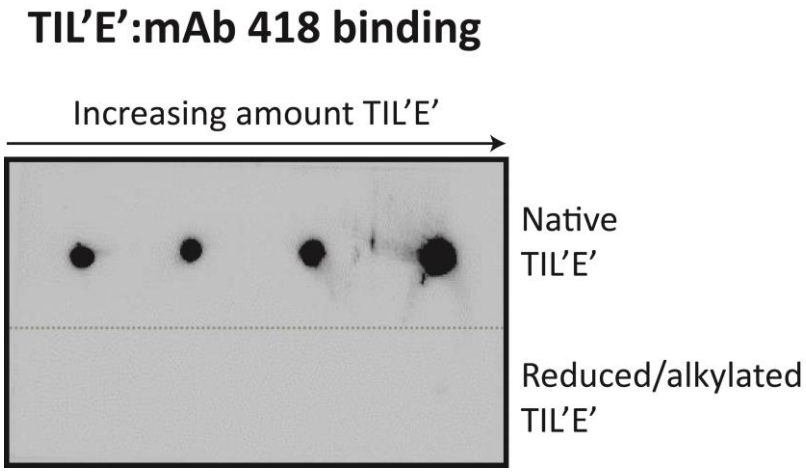


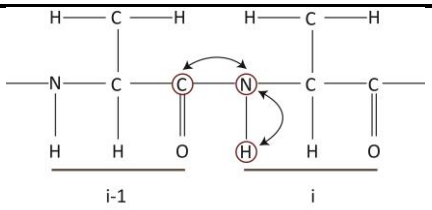
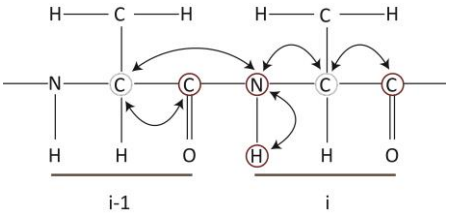
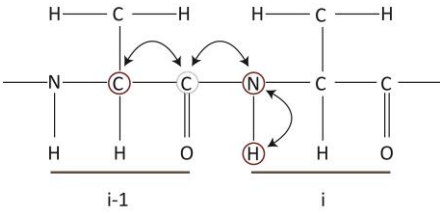
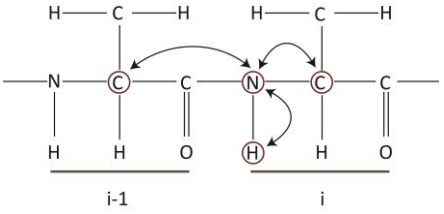
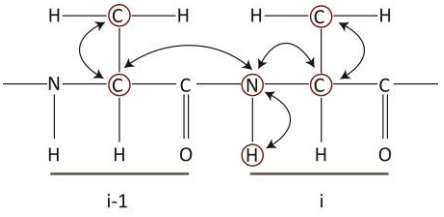
Figure 2.13: Dot blot analysis of TIL'E' functionality – binding to mAb 418. Cleaved TIL'E' (0.45 mg/mL) was reduced with 10 mM dithiothreitol and the TIL'E' thiol groups were subsequently blocked with 30 mM iodoacetamide to prevent their re-oxidation. Native, and reduced and alkylated TIL'E' were spotted onto the nitrocellulose membrane in increasing amounts of 56.25 ng, 112.5 ng, 225 ng and 450 ng. The membrane was incubated with mAb 418 in blocking buffer and binding of mAb 418 to TIL'E' was revealed by ECL detection of secondary HRP-conjugated anti-mouse antibody. mAb 418 binds to native TIL'E' but not to reduced and alkylated TIL'E'.

That TIL'E' binds to both FVIII and mAb 418 is clearly attested to by dot blotting studies, and importantly, reduced and alkylated TIL'E' does not bind mAb 418 (Figure 2.13). Together, these results reveal a specific TIL'E'·FVIII interaction that is not dependent upon glycosylation, and affirms that the recombinantly expressed TIL'E' is a faithful facsimile of these domains in full-length VWF - adopting a biologically active native fold, complete with the correct pattern of disulphide bridges. Importantly, binding of TIL'E' to FVIII identifies the minimal VWF domains that form the major FVIII binding region and indicate that TIL'E' may be sufficient to stabilise FVIII survival *in vivo*.

2.5. Resonance assignment of TIL'E' backbone atoms

Once it had been established that the recombinantly produced TIL'E' is both correctly folded and functional, the chemical shifts of the TIL'E' nuclei were assigned by standard triple-resonance experiments (Table 2.2).

Table 2.2: Triple resonance experiments used for sequential backbone assignment[342].

Experiment	Correlations observed	Magnetisation transfer
HNCO	$^{1}H^{N}_{i} - {}^{15}N_{i} - {}^{13}CO_{i-1}$	
HN(CA)CO	$^{1}H^{N}_{i} - {}^{15}N_{i} - {}^{13}CO_{i}$ $^{1}H^{N}_{i} - {}^{15}N_{i} - {}^{13}CO_{i-1}$	
HN(CO)CA	$^{1}H^{N}_{i} - {}^{15}N_{i} - {}^{13}C\alpha_{i-1}$	
HNCA	$^{1}H^{N}_{i} - {}^{15}N_{i} - {}^{13}C\alpha_{i}$ $^{1}H^{N}_{i} - {}^{15}N_{i} - {}^{13}C\alpha_{i-1}$	
HNCACB	$^{13}C\beta_{i}/{}^{13}C\alpha_{i} - {}^{15}N_{i} - {}^{1}H^{N}_{i}$ $^{13}C\beta_{i-1}/{}^{13}C\alpha_{i-1} - {}^{15}N_{i} - {}^{1}H^{N}_{i}$	

The experiments outlined in Table 2.2 are used in the initial stage of any protein investigation by NMR spectroscopy, to make connections between ^{1}H, ^{13}C and ^{15}N nuclei. The patterns of connections are then mapped onto the protein sequence to assign all resonances belonging to each amino acid residue. The nuclear spins attached

to each of the TIL'E' ^{1}H-^{15}N resonances observed in the ^{15}N-HSQC (Figure 2.9) were subsequently identified in the HNCO, HN(CA)CO, HN(CO)CA, HNCA and HNCACB spectra, to obtain an assignment of TIL'E' backbone atoms (Table 2.2). These standard triple-resonance experiments correlate each ^{1}H-^{15}N amide group with the ^{13}CO, ^{13}Cα and ^{13}Cβ of its own residue (i), and with that of the preceding residue (i–1), thereby permitting the assignment of individual spin systems (nuclear spins from a single residue), which are then correlated to a specific amino acid in the TIL'E' sequence (Figure 2.14).

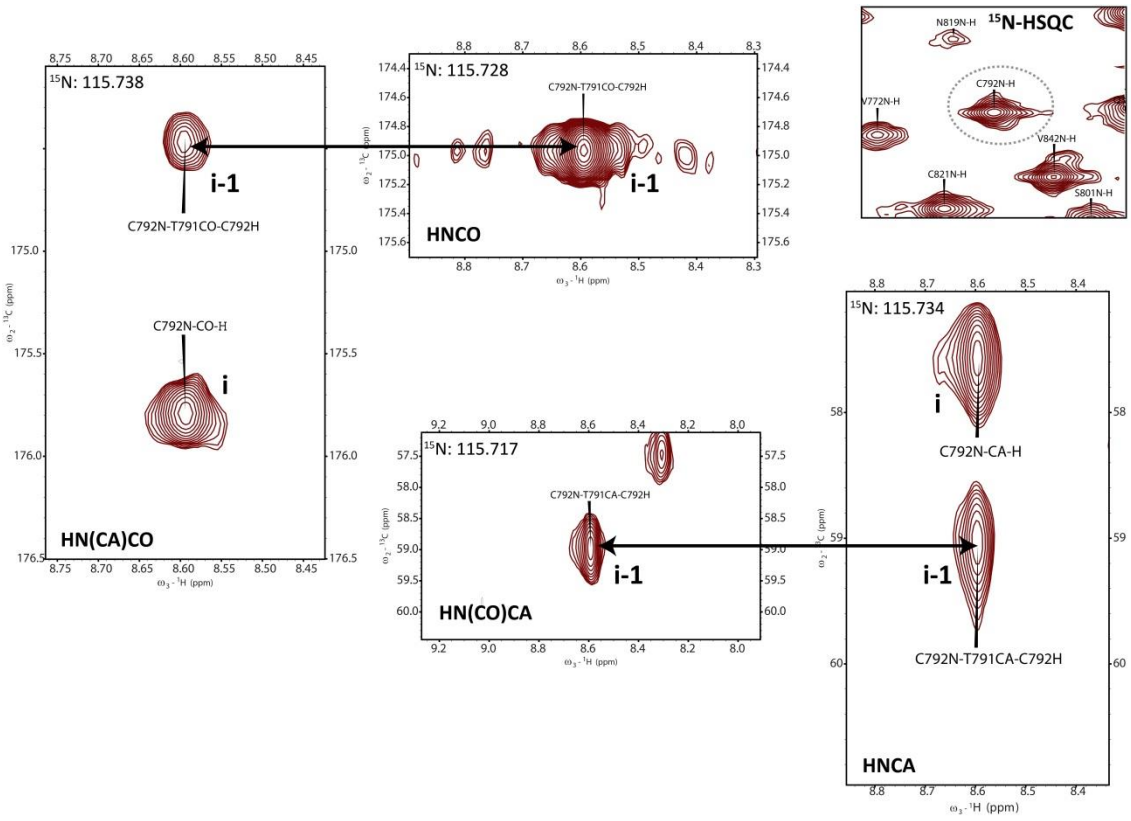


Figure 2.14: Correlation of a single spin system. Cys792 ^{1}H-^{15}N (designated i) resonance from the assigned ^{15}N-HSQC spectrum is linked to its i-1 (Thr791) ^{13}CO resonance in the HNCO and then to its own CO shift in the HN(CA)CO spectrum. The Cys792 Cα and Thr791 Cα are similarly linked in the HN(CO)CA and HNCA spectra as shown. Thus, the CO_i, CO_{i-1}, $C\alpha_i$ and $C\alpha_{i-1}$ resonances of Cys792 are known. The intra-residue peak is often of greater intensity than the inter-residue peak, aiding assignment.

Identification of the nuclear spins attached to each ^{1}H-^{15}N resonance in the ^{15}N-HSQC spectrum facilitates their sequential connectivity, such that the carbon$_i$ (^{13}CO, ^{13}Cα and ^{13}Cβ) chemical shift of one residue is aligned to the carbon$_{i-1}$ (^{13}CO, ^{13}Cα and ^{13}Cβ) shift of the following residue in the TIL'E' sequence, and henceforth (Figure 2.15). These

patterns of connections were subsequently mapped onto the TIL'E' amino acid sequence to specifically assign a spin system to each residue. This procedure is greatly assisted by the distinct Cα and Cβ chemical shifts of glycine (Cα~45.38), isoleucine (Cα~61.62 ppm), proline (Cα~63.32 ppm), valine (Cα~62.53 ppm), alanine (Cβ~19.07 ppm), serine (Cβ~63.72 ppm) and threonine (Cβ~69.61 ppm)[343].

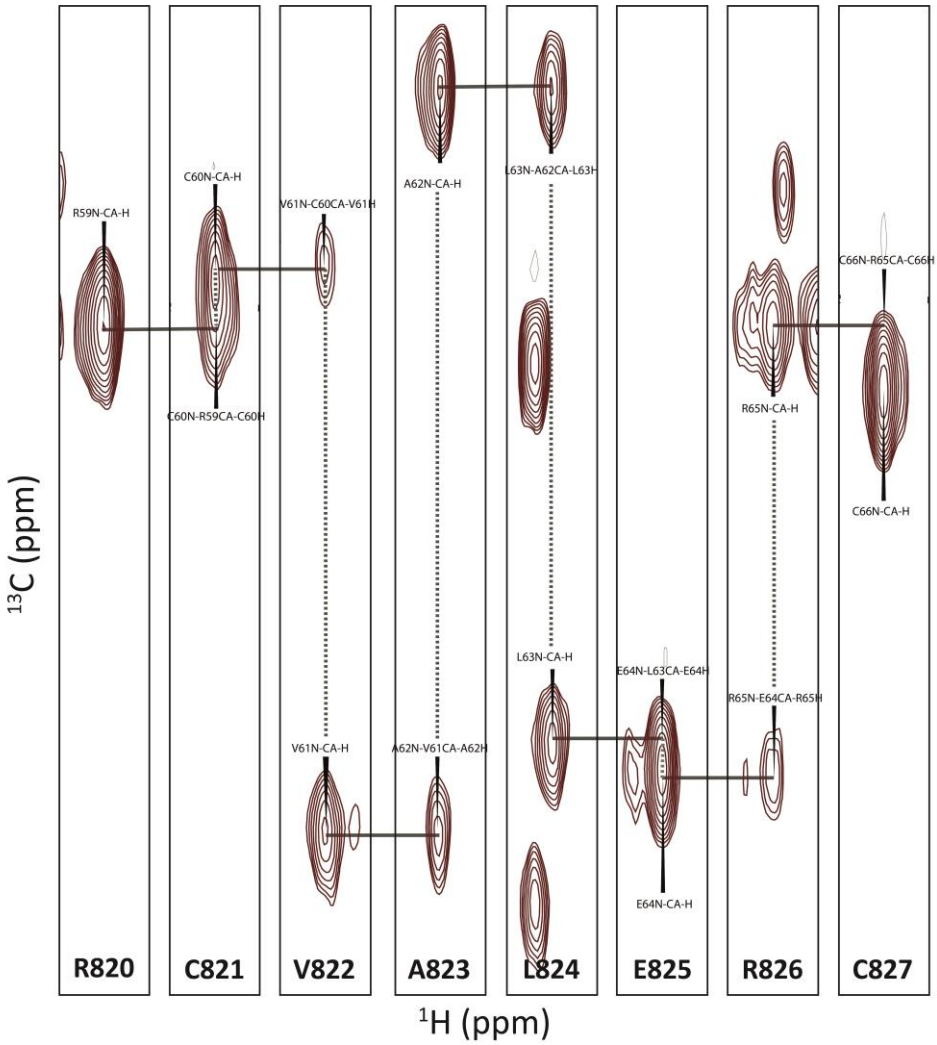


Figure 2.15: Sequential TIL'E' backbone walk. Shown are strips from the 3D HNCACB spectrum, illustrating the sequential linkage of $^{13}C\alpha$ atoms. The connectivity from $C\alpha_i$ to $C\alpha_{i-1}$ to $C\alpha_i$ in the preceding residue's strip are indicated to sequentially link spin systems.

In this manner, 78% of the TIL'E' $^{1}H^{N}$, ^{13}C and ^{15}N backbone nuclei were assigned, whereby the missing resonances are contributed to by seven proline residues (Pro769, Pro770, Pro777, Pro811, Pro812, Pro828 and Pro838), and the remaining unassigned nuclei reside predominantly in the TIL' region between residues 781 and 805 (Figure 2.16).

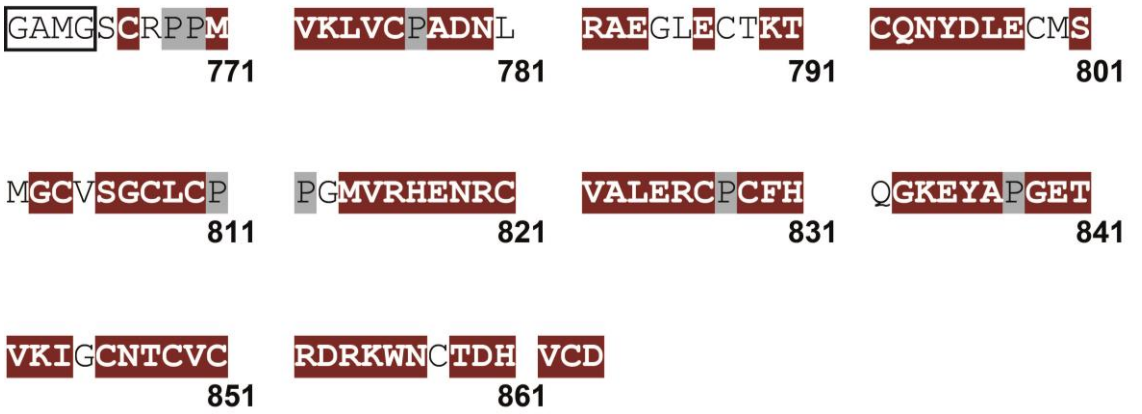


Figure 2.16: Assigned TIL'E' ^{1}H, ^{13}C and ^{15}N backbone nuclei. The TIL'E' construct is composed of 103 residues, of which the N-terminal four amino acids, shown boxed, constitute cloning artefacts (GAMG). The VWF sequence begins at the fifth residue corresponding to Ser766 and ends at Asp864. The backbone $^{1}H^{N}$, ^{13}C (CO, Cα Cβ) and ^{15}N chemical shifts of VWF residues that have been assigned are highlighted in red. Proline residues are highlighted in grey. Unassigned residues are predominantly found between residues 781 and 805 of domain TIL'.

The assigned ^{15}N-HSQC spectrum (Figure 2.17, upper panel) enabled a more detailed analysis of the peak intensity inhomogeneity that was observed on initial inspection following TIL'E' HSQC spectral acquisition. It appears that the effect of peak broadening is primarily confined to resonances assigned to domain TIL', in particular at positions Cys767, Ala778, Asn780, Ala783, Cys810, Asn819, Gly833 and Cys846 (Figure 2.17, lower panel). This indicates that TIL' is undergoing chemical exchange between different conformations, and is consistent with missing resonances in the ^{15}N-HSQC spectrum (79 peaks observed of the 93 expected) and the location of unassigned backbone residues in the TIL'E' sequence (Figure 2.16).

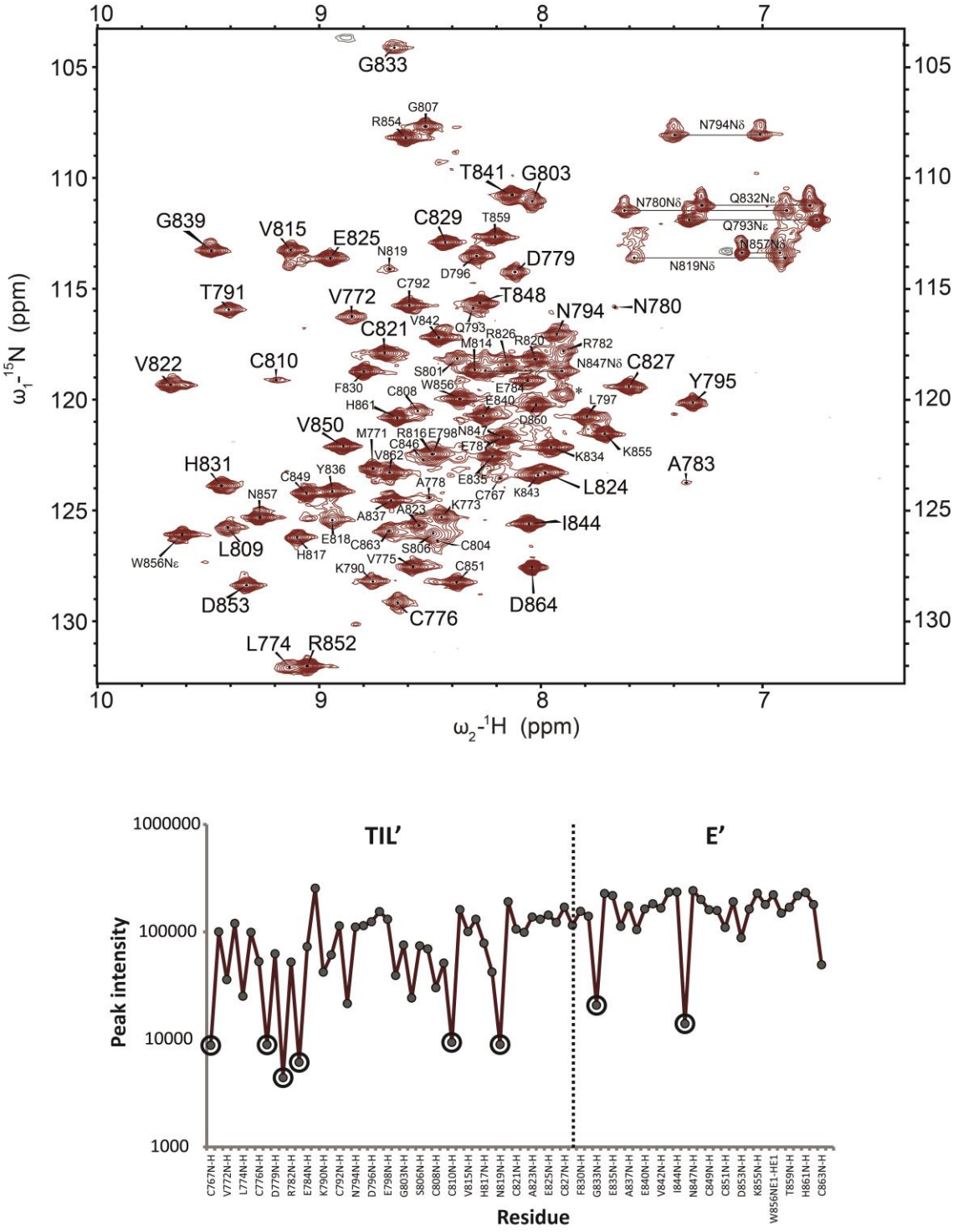


Figure 2.17: Assigned ^{15}N-HSQC NMR spectrum of $^{13}C^{15}$N TIL'E' and peak intensity analysis. The assigned ^{15}N-HSQC spectrum recorded at 600 MHz in 20 mM sodium phosphate, 100 mM NaCl, pH 7.4, 25°C is shown in the upper panel. A near complete backbone assignment of ^{1}H-^{15}N resonances is achieved. Of the 79 peaks observed, 77 backbone amides and the indole of Trp856 are assigned. A single unassigned peak in the centre of the spectrum is indicated by *. Side chain NH_2 groups of Asn and Gln residues are assigned and connected with grey bars. There are broad peaks with lower intensity seen in the spectrum, the plot in the lower panel confines the peak intensity heterogeneity to domain TIL', in particular to residues Cys767, Ala778, Asn780, Ala783, Cys810, Asn819, Gly833 and Cys846 (circled) and this is highly suggestive of conformational exchange in at least one region of TIL'. The intense C-terminal peak of Asp864 has been omitted from the plot for clarity.

2.6. TIL'E' backbone chemical shifts predict secondary structure, dynamics and oxidation state of cysteine residues

Assignment of TIL'E' backbone atoms permitted a preliminary assessment of the secondary structure elements and conformational flexibility of TIL'E'. NMR chemical shifts of backbone nuclei depend on the secondary structure of the individual residues. This allows accurate mapping of protein backbone order based on the observation that random coil chemical shifts are characteristic of highly flexible regions, whereas non-random coil shifts are commonly found in rigid areas of protein structures. Therefore, the TIL'E' backbone shifts were used to provide an initial indication of TIL'E' secondary structure and of backbone flexibility, as given by the random coil index (RCI) derived order parameter, S^2, which describes the amount of local mobility. Extreme RCI-S^2 values of 1 and 0 correspond to fully restricted and unconstrained internal motion of the backbone, respectively[344–346] (Figure 2.18).

TIL' is less structured and more dynamic than domain E'. The unassigned TIL'E' backbone nuclei (residues 781 to 805) cluster in unstructured loop regions, to the greatest extent between predicted strands β1 and β2 of domain TIL' (Figure 2.18, upper panel). This region is coincident with low RCI-derived order parameters indicating a dynamically disordered loop, that is significantly more flexible than the remainder of TIL'E' (Figure 2.18). The peak broadening, observed for residues Cys767, Ala778, Asn780, Ala783, Cys810, Asn818 in TIL' and Gly833 and Cys846 in E' (Figure 2.17), is accounted for by their position in predicted loop regions, of which Ala778, Asn780 and Ala783, occur in the β1-to-β2 loop of domain TIL' (Figure 2.18, lower panel). Furthermore, broader peaks were largely found in domain TIL' and this is consistent with this domain exhibiting a greater degree of backbone flexibility.

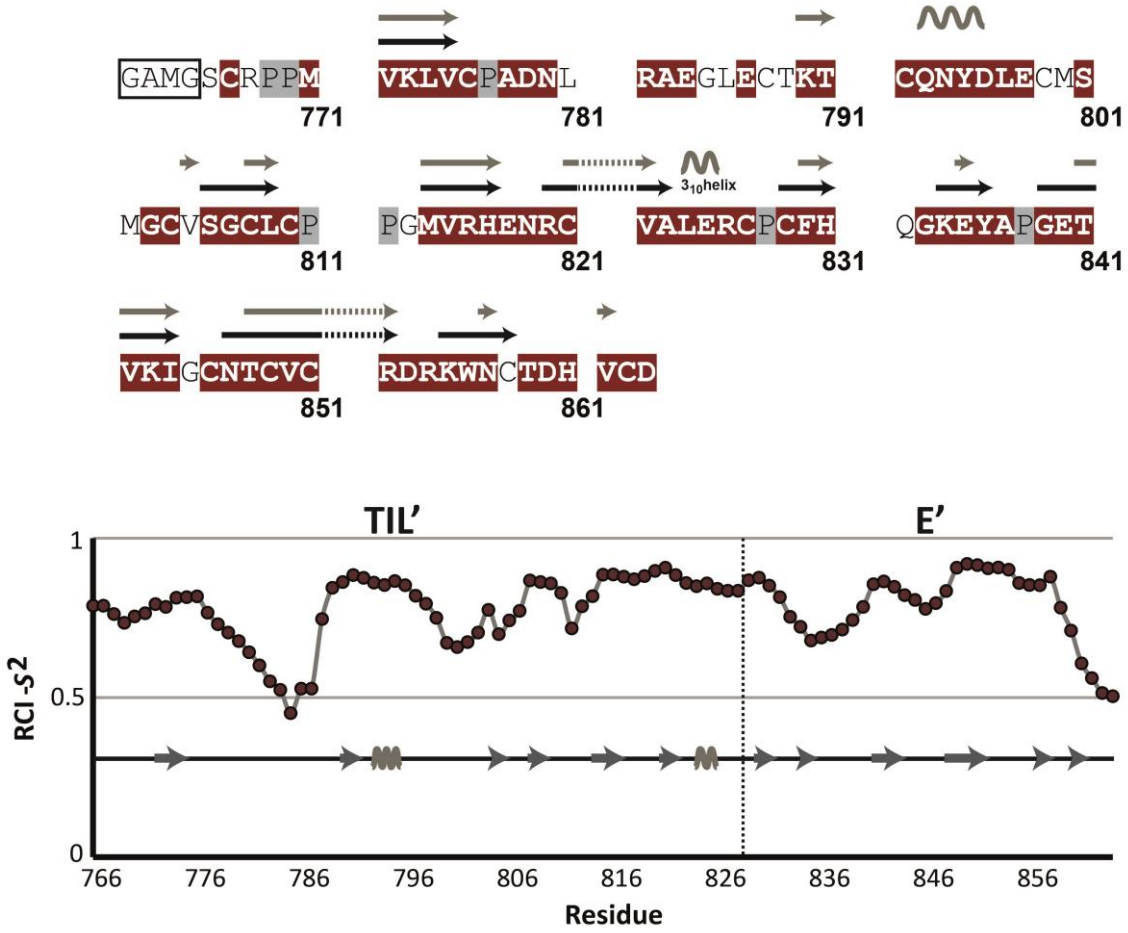


Figure 2.18: Secondary structure and order parameter predictions of TIL'E'. The backbone ^{1}H, ^{13}C and ^{15}N chemical shifts that have been assigned are highlighted in red. Proline residues are highlighted in grey. The secondary structure of TIL'E' is shown above the sequence; in grey are the predicted secondary structure elements (arrow, β-strand; wave, α-helix) from the backbone chemical shifts, and in black are shown the experimentally derived secondary structure elements from the high resolution solution structure of TIL'E'. Unassigned residues and residues that exhibit peak broadening are found in loop regions. The bottom panel shows the random coil index (RCI) derived order parameter (S^2) prediction[345], and in the plot is shown the TIL'E' predicted secondary structure for clarity. An RCI-S^2 value of 0 indicates completely unrestricted motion, whereas an RCI-S^2 value of 1 indicates rigidity. Dynamic residues are taken to be those with an order parameter value less than 0.6. The loop between predicted strands β1 and β2 of the TIL' domain is significantly more dynamic than the remainder of the molecule.

In addition to secondary structure and backbone flexibility predictions, the backbone chemical shifts were used to ascertain the redox state of cysteine residues. Cβ chemical shifts of the assigned Cys residues provided further evidence that the sixteen cysteine residues of TIL'E' are oxidised and engaged in disulphide bonds. A database of cysteine ^{13}C Cα and Cβ shifts, constructed from published NMR structures, revealed

that reduced and oxidised cysteines have characteristic chemical shifts[347]. The $^{13}C\alpha$ chemical shifts of reduced and oxidised cysteine range narrowly from 50.5 ppm to 65.6 ppm, and hence, exhibit considerable overlap; whereas, the $^{13}C\beta$ shifts of reduced and oxidised cysteine span a wider range (23.8 to 50.9 ppm), and clearly fall into two distinct clusters[347] (Figure 2.19). Therefore, $^{13}C\beta$ chemical shifts are diagnostic of disulphide bond formation and can discriminate between cysteine residues in the reduced and oxidised state. The observed Cβ shift for oxidised cysteine is 40.7 $\pm$ 3.8 ppm, and for reduced cysteine is 28.4 $\pm$ 2.4 ppm, and is thus, capable of defining the Cys residue redox state (Reduced $<$ 32.0 ppm; oxidised $>$ 35.0 ppm)[347].

The Cβ shifts of all cysteines assigned in TIL'E' (Cys776, 40.1 ppm; Cys792, 36.3 ppm; Cys804, 38.6 ppm; Cys808, 41.3 ppm; Cys810, 36.6 ppm; Cys821, 40.9 ppm; Cys827, 37.2 ppm; Cys829, 42.4 ppm; Cys846, 37.8 ppm; Cys849, 50.5 ppm; Cys851, 36.9 ppm; Cys858, 47.9 ppm; Cys863, 37.6 ppm) are indicative of oxidised Cys residues (Figure 2.19). Of the remaining three cysteines, the signal from N-terminally placed Cys767 is very weak, which prevented assignment of its Cα/Cβ shifts; Cys788 and Cys799 are predicted to participate in a 3-5 linkage in an unstructured region, and remain unassigned (Figure 2.18). The Cβ shifts of cysteine residues may also be correlated with secondary structure, such that the mean chemical shift for an oxidised cysteine in a β-strand is 43.0 ppm (range: 35.9-50.9 ppm; median: 43.2 ppm)[347]; the chemical shifts of cysteines Cys808, Cys821, Cys829, Cys849, Cys851 and Cys858, support their predicted secondary structure assignment to β-strands (Figure 2.18, upper panel).

In summary, 78% of TIL'E' backbone atoms have been assigned. Unassigned residues and residues exhibiting peak broadening are predominately localised to flexible loop regions in domain TIL'. Importantly, the full formation of eight disulphide bonds in TIL'E' is further supported by their Cys residue $^{13}C\beta$ chemical shifts.

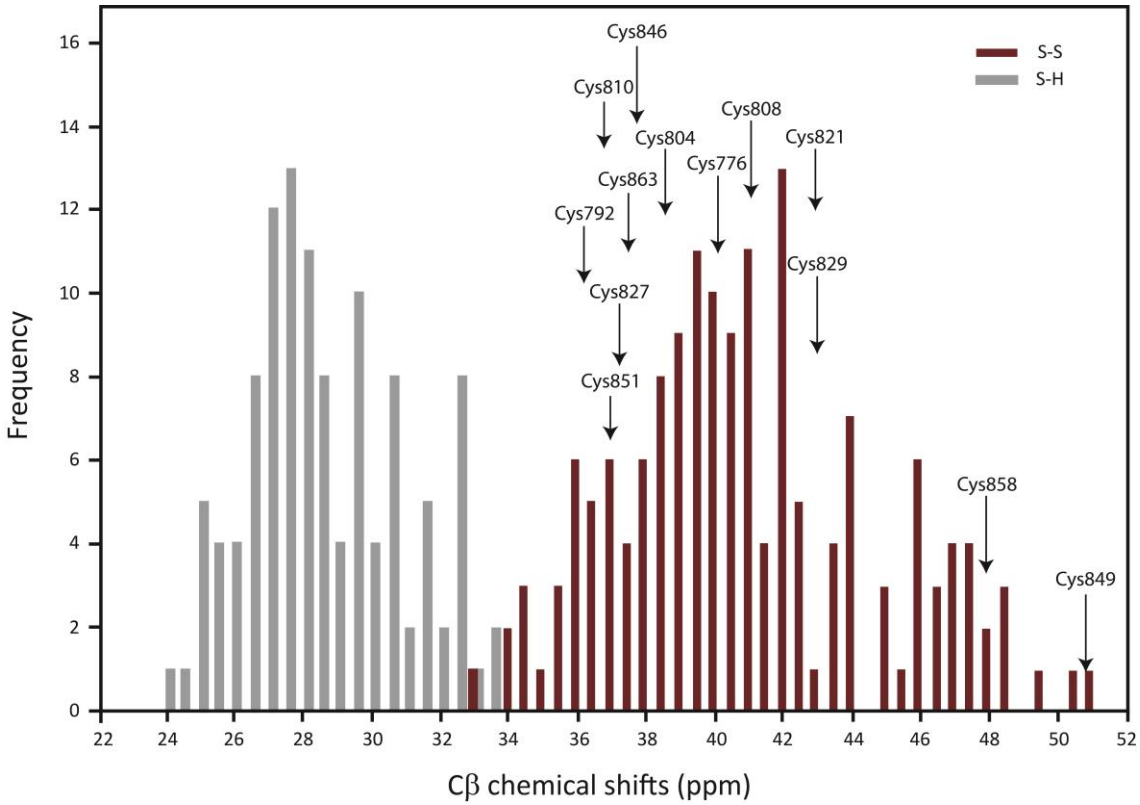


Figure 2.19: Distribution of cysteine Cβ chemical shifts as a function of redox state. Shown is the distribution plot of Cβ chemical shifts, as derived by Sharma *et al*[347]. The Cβ chemical shifts of reduced (grey bars) and oxidised (red bars) cysteine fall in two distinct domains, but overlap in a narrow range of 33.0 ppm to 34.0 ppm[347]. These shifts however, are also clustered as a function of secondary structure. The mean Cβ shifts of oxidised and reduced cysteine in a β-sheet are 43.0 ppm and 29.7 ppm, respectively. The positions of TIL'E' cysteine Cβ shifts, in relation to this distribution plot, are shown. They are strongly indicative of oxidised cysteine residues.

2.7. Assignment of TIL'E' side chain atoms

The accuracy of the calculated NMR structure is dependent upon a near complete assignment of all atoms in the TIL'E' molecule; this ensures that the subsequently derived structural restraints reflect maximal coverage of the molecule. Accordingly, the resonances of aliphatic and aromatic side chain ^{1}H and ^{13}C nuclei were assigned using TOCSY (total correlation)-based experiments, CC(CO)NH-TOCSY, HCC(CO)NH-TOCSY, HCCH-TOCSY, which correlate every intra-residue ^{1}H or ^{13}C distal to the α position down the amino acid chain (Figure 2.20).

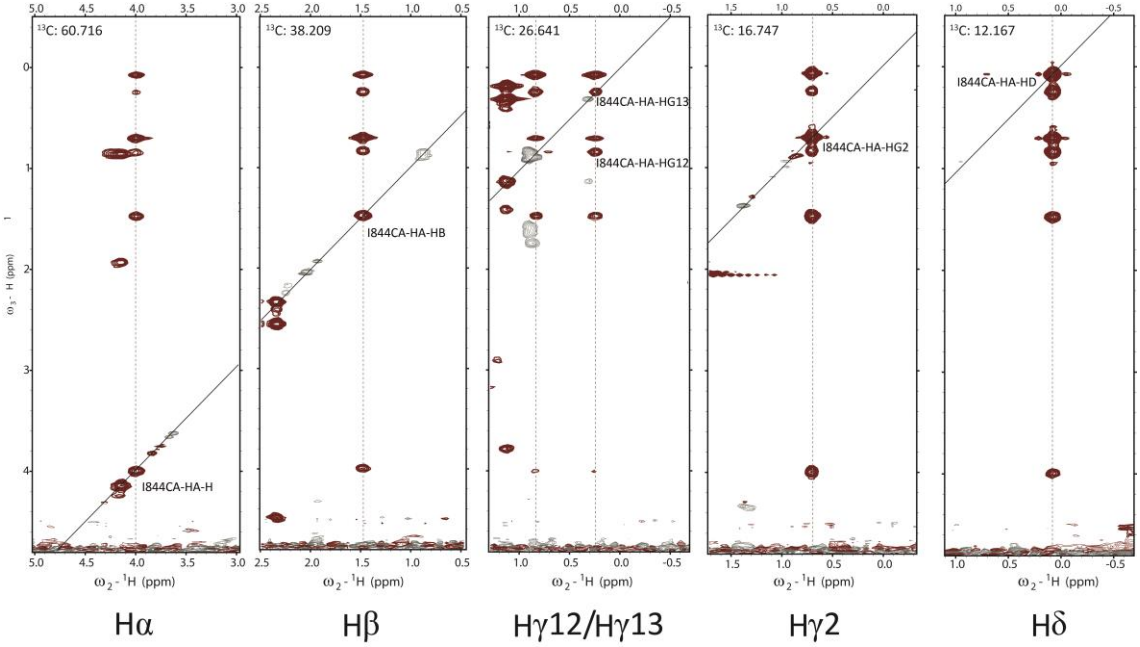


Figure 2.20: 3D HCCH-TOCSY assignment of Ile844 side chain protons. In the HCCH-TOCSY experiment magnetisation is transferred from the side chain hydrogen nuclei to their attached ^{13}C nuclei. This is followed by isotropic ^{13}C mixing and finally transfer back to the side chain hydrogen atoms for detection. This spectrum yields strips at each carbon frequency in the side chain in which all side chain hydrogen (proton) resonances are visible. Thus, in each CH strip, all the hydrogen atoms attached to all other carbon atoms in the same chain are observed, as shown for Ile844.

The amide side chains of Asn and Gln residues, methyl groups of Met and the diastereotopic methyl groups of Val and Leu, were also assigned. Assignment of Val and Leu methyl groups greatly improves the precision and accuracy of structure determination, as these groups are involved in numerous identifiable inter-residue and long range contacts[348], and were used as restraints in the subsequent structure calculation of TIL'E'. The unambiguous assignment of prochiral methyl groups in valine and leucine, by the elegant method of fractional labelling[349], is based on the premise that biosynthesis of these residues from glucose is stereoselective. Biosynthetic fractionally labelled TIL'E' was obtained by growing *E. coli* in minimal media containing 10% ^{13}C-D-glucose and 90% unlabelled glucose as the sole carbon source. The carbon positions in these preparations are ^{13}C labelled to an extent of 10%. Disregarding the natural ^{13}C abundance of 1.1% in the unlabelled glucose, the probability that two adjacent carbon positions are labelled in the same molecule is then 1%, unless the two carbon atoms originate from the same carbon source molecule so that this probability becomes 10%[349]. These two distinct situations prevail for the isopropyl group of valine and leucine (Figure 2.21).

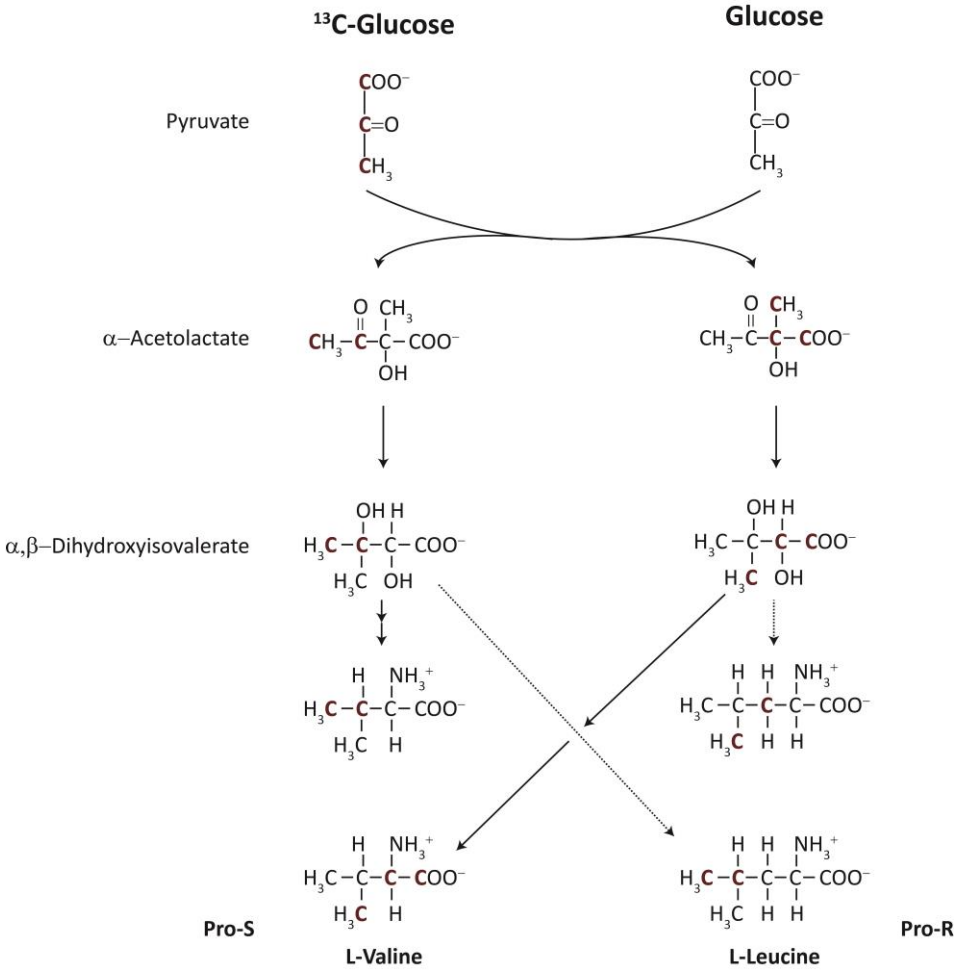


Figure 2.21: Stereoselective biosynthesis of valine and leucine. Reaction pathways for the biosynthesis of valine and leucine from a mixture of fully ^{13}C-labelled and unlabelled glucose showing the stereochemistry and the principal labelling patterns[349]. This methyl migration is stereoselective, and the migrating methyl group becomes the Pro-S in valine and leucine, γ^2CH3 and δ^2CH3, respectively[349]. A red carbon indicates a ^{13}C label, and a black carbon indicates the natural ^{13}C abundance of 1.1%.

The Pro-R methyl carbon (Cγ1 in Val and Cδ1 in Leu) derives from the same pyruvate as the adjacent branching methine carbon (-CH) and are, in the absence of isotope scrambling, labelled with ^{13}C in the same molecule, while the Pro-S methyl carbon (Cγ2 in Val and the Cδ2 in Leu) and the adjacent carbon originate from two different pyruvate molecules (Figure 2.21). Therefore, if the pro-S methyl group is enriched with ^{13}C, there is a probability of only 1% that the adjacent –CH group in the same molecule is also labelled. Consequently, the valine and leucine methyl groups of fractionally labelled TIL'E' were readily identified in a constant time ^{13}C-HSQC spectrum, in which

the Cγ1 and Cδ1 methyl group (pro-R) peaks appear with opposite sign (-ve), as they have an even number of directly bonded ^{13}C atoms, to the Cγ2 and Cδ2 methyl groups (pro-S).

A total TIL'E' atom assignment of 88.9% was achieved (Figure 2.22). Unassigned atoms correspond to residues in unstructured regions of TIL'E', with the exception of Lys773 in domain TIL', and Phe830, Tyr836 and Cys858 in E' that are in β-strands (Table 2.3).

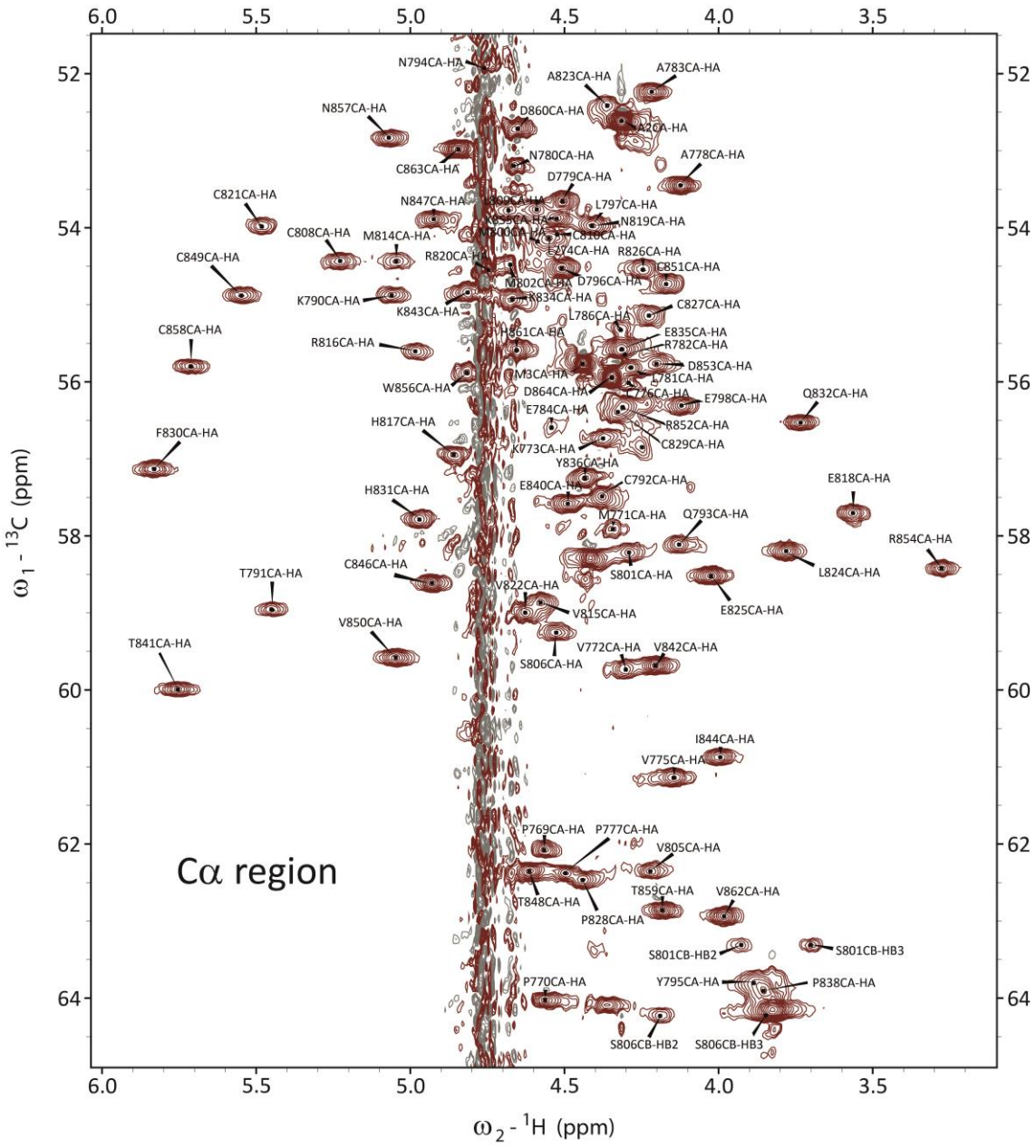


Figure 2.22: CT ^{13}C-HSQC spectrum showing the ^{13}Cα-^{1}Hα chemical shift assignments of TIL'E'. This spectrum is the carbon equivalent of the ^{15}N-HSQC spectrum. In the constant time (CT) ^{13}C-HSQC all ^{13}C-^{1}H correlations are observed. The assigned Cα region is shown and 88.9% of all TIL'E' resonances have been assigned. Peaks appear with opposite sign to the ^{13}Cα-^{1}Hα peaks if they have an even number of directly bonded aliphatic carbons.

Table 2.3: Unassigned TIL'E' protons. Q denotes diastereotopic groups.

Amino acid	Unassigned protons
Ser766	H^N Hα Qβ
Cys767	Hα Qβ
Arg768	H^N Hα Qβ Qγ Qδ
Pro769	Qγ
Lys773	Qδ Qε
Leu781	H^N Hγ
Gly785	H^N Qα
Leu786	H^N
Glu787	Hα Qβ Qγ
Cys788	H^N Hα Qβ
Thr789	H^N Hα Qβ
Lys790	Qδ
Tyr795	Hε2 Hδ2
Cys799	H^N Hα Qβ
Met800	H^N Qε
Met802	H^N Qβ Qε
Val805	H^N
Pro811	Hα Qβ Qγ Qδ
Pro812	Hα Qβ Qγ Qδ
Gly813	H^N
Phe830	Hε2 Hδ2
Gln832	H^N
Tyr836	Hε2 Hδ2
Gly845	H^N
Cys858	H^N

2.8. NOE distance restraints

Following extensive assignment of nearly all backbone and side chain resonances in TIL'E' (Appendix 5.2), 108 backbone dihedral angle restraints (ϕ and ψ) were calculated from ^{15}N, $^{1}H^N$, ^{13}CO, $^{13}C\alpha$, and $^{13}C\beta$ backbone chemical shifts using the programme TALOS+[345] and NOE (nuclear Overhauser enhancement)-based experiments for the extraction of distance restraints were analysed. The most important NMR-observable parameter used in determining protein structure is the NOE. Unlike the triple-resonance experiments used to assign the TIL'E' backbone chemical shifts, whereby magnetisation is transferred from one nucleus to another directly through bonds connecting atoms, the NOE is a measurement of 'through-space' interactions between atoms that are spatially close. The NOE peak intensity is proportional to the inverse sixth power of the distance between two interacting ^{1}H nuclei, hence the r^{-6}

dependence permits the detection of NOEs only between protons within a 6 Å distance[342]. ^{13}C- and ^{15}N-edited NOESY experiments were used for the measurement of NOEs, which are correlated with inter-nuclear ^{1}H-^{1}H distances of assigned resonances, and in turn, used as distance constraints in structural calculations to determine the TIL'E' fold (Figure 2.23).

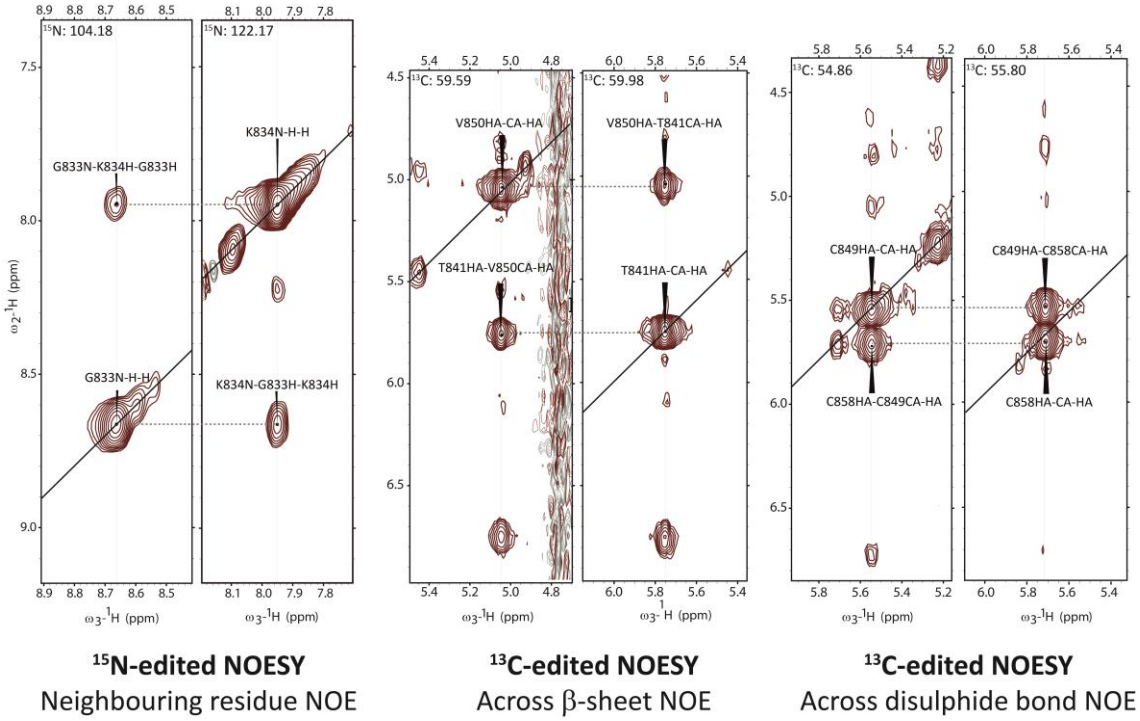


Figure 2.23: NOEs as structural restraints. Shown are NOESY spectra strips from ^{15}N-edited NOESY and ^{13}C-edited NOESY experiments. Each strip in a ^{15}N-edited NOESY contains NOEs from one ^{1}H-^{15}N amide group to all other hydrogen atoms in its proximity, whereas in the ^{13}C-edited NOESY, each strip contains NOEs from one –CH group to all other hydrogens nearby. Examples of assigned NOE peaks to a neighbouring residue (Gly833 to Lys844; E'), across a β-sheet (Thr841 to Val850; E' β3 to β4) and across disulphide bond NOE (Cys849 Hα to Cys858 Hα; E' 3-5 linkage) are given.

A total of 954 unique inter-residue distance restraints were unambiguously identified using an iterative strategy, in which successive cycles of structure calculations included an increasing number of inter-residue NOE restraints; this method permitted the resolution of NOE ambiguities and assignment of additional NOEs by reference to a preliminary TIL'E' structure model. NOE restraints that were not collaterally supported by other NOEs to residues in proximal vicinity, were discarded to avoid propagating errors that may influence assignment of other NOEs, or worse, restrain an incorrect fold. The steps from assignment to structure and error analysis were performed in a repetitious manner until an optimised data set was obtained, from which the final

TIL'E' structure models were calculated. Of the unique NOE distance restraints used in the final structure calculation, all of which were assigned in all three dimensions, 353 are long-range, distances between hydrogen nuclei that are more than five residues apart in the primary sequence; 97 are medium range, sequence separation of four to five residues; 121 are short-range, sequence separation of two to three residues; and 383 NOE restraints are between sequential residues. NOEs obtained for each residue are summarised in Figure 2.24. Overall there is good NOE coverage of the TIL'E' molecule; however, there is a greater degree of NOE density in domain E' that is most likely due to both the localisation of unassigned residues to TIL' and its predicted flexibility, giving rise to broader and weaker signals that cannot be unambiguously assigned.

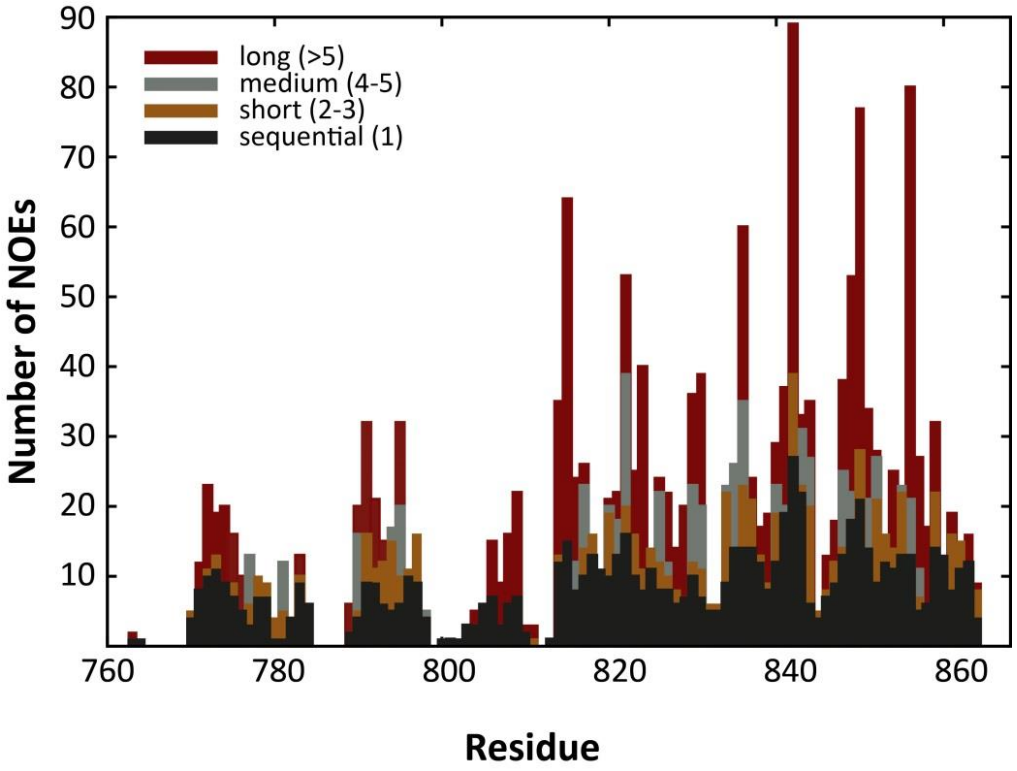


Figure 2.24: Assigned NOEs per residue of TIL'E'. The number of sequential NOEs (black), short-range NOEs (brown; sequence separation of two to three residues), medium range NOEs (grey; sequence separation of four to five residues) and long range NOEs (red; sequence separation of more than five residues) are plotted against TIL'E' residues. There is a greater degree of NOE density in domain E' (aa 829-863) than in TIL' (aa 767-827).

2.9. NOE-based confirmation of TIL'E' disulphide bond topology

Formation of a disulphide bond demands the close proximity of sulphur atoms (2.03 Å), an angle between the disulphide and the β-carbon (Cβ-Sγ) that is close to 103°, and that the two β-carbons are orientated to correspond to a rotation close to either +90° or −90° about the disulphide bond[336,350]. A disulphide bond is the only molecular interaction that imposes such stringent stereochemical requirements on a protein, and with respect to structure determination in solution state NMR spectroscopy, the correct identification of disulphide bonds constitutes a powerful conformational constraint, that considerably restricts the structural space available[336]. The NMR active isotope of sulphur (^{33}S) has highly unfavourable magnetic properties[351], making the unequivocal establishment of disulphide bond connectivity by NMR alone difficult, particularly in unstructured regions of a protein; however, the multipronged approach used here, supported by biochemical data, has obviated this difficulty.

The designations of NOESY cross peaks between side chain β-protons of two Cys residues were used to infer that they are covalently connected by a disulphide linkage. This is a reasonable assumption, since the distance between the Cβ atoms of a bonded Cys pair ranges from 3.5 Å to 4.3 Å, with a mean of 3.8 Å[352–354], and consequently the inter-residue distance between the pair of β-protons in a disulphide bond will always be less than the NOE upper limit (~6 Å)[351]. The mean distance between the Cα atoms of Cys residues in a disulphide bond is 5.63 Å[321,354,355], and approaches the limit for NOE detection, however some NOEs are seen between the α-protons of Cys residues in a disulphide bond, such as between Cys849-Cys858 (Figure 2.23 and Table 2.4). Up to four β-proton NOE peaks may be observed between the two Cys residues in a pair, and up to four NOEs between the α-proton of one Cys and the β-proton of its partner Cys residue (for both α carbons). Five disulphide bonded cysteines were investigated and at least four NOE cross peaks were found across each linkage; furthermore, there were no 'antidiagnostic' NOEs between Cys H^N and Cys Hα of its partner residue, which would strongly indicate the absence of a disulphide bond[356] (Figure 2.25 and Table 2.4).

Table 2.4: Summary of NOEs observed between coupled cysteines. ✓ indicates a unique NOE.

Disulphide bond	Domain connectivity	Hα-Hα	Hα-Hβ	Hβ-Hβ	H^N-Hα
767-808	TIL': 1-7	-	-	-	-
776-804	TIL': 2-6			✓✓✓✓	
788-799	TIL': 3-5	-	-	-	-
792-827	TIL': 4-10			✓✓✓✓	
810-821	TIL': 8-9	-	-	-	-
829-581	E': 1-4		✓✓✓	✓✓✓✓	
846-863	E': 2-6		✓✓✓	✓✓	
849-858	E': 3-5	✓	✓✓✓✓	✓✓	

Of the three remaining TIL' Cys pairings (Cys767-Cys808, Cys788-Cys799 and Cys810-Cys821): the N-terminal Cys767 (1-7 linkage) and cysteines 788 and 799 (3-5 linkage) are unassigned (Table 2.3). Lastly, no observable NOEs were found across the assigned Cys810-Cys821 residues (8-9 linkage). A disulphide bonding pattern other than that which was predicted, is excluded by the absence of NOEs between non-paired cysteine residues, supporting the unequivocal inference that the TIL'E' model is of the correct fold. It is important to note, that six NOE peaks across the homology-predicted Cys846-Cys863 disulphide bond in domain E' confirms their covalent linkage in a 2-6 configuration (Figure 2.25). A disulphide bond is composed of six atoms linking the two α-carbon atoms of the cysteine residues; Cα-Cβ-Sγ-Sγ'-Cβ'-Cα'. These atoms define five χ angles that can be either positive or negative, resulting in twenty possible disulphide bond configurations[355]. However, because disulphide bonds exist in different conformations in the various NMR models[357], this precluded their finer characterisation in TIL'E', based upon bond geometry[355,358].

Further confirmation of the validity of the Cys pairings was afforded by the observation that structure calculations performed with inclusion and exclusion of the eight disulphide bond constraints was achieved without perturbation to the global TIL'E' fold. In particular, when disulphide bond constraints were excluded, all predicted cysteine pairs lay in close proximity in the TIL'E' structure, with the exception of bond Cys788-Cys799 (3-5 linkage) and to some extent bond Cys776-Cys804 (2-6 linkage) in domain TIL' β1-to-β2 loop, that were heterogeneous and poorly defined. Furthermore, all Cβ atoms were closer to the Cβ atom of a partner Cys residue than to another Cys

Cβ in spatial proximity, thereby eliminating connectivity ambiguities for disulphide bonds Cys792-Cys827 (linkage 4-10 of TIL') and all three Cys-Cys pairings in E', that have been predicted but not biochemically assigned. In accordance with the preceding observations, all disulphide bonds were included as constraints in subsequent structure calculations.

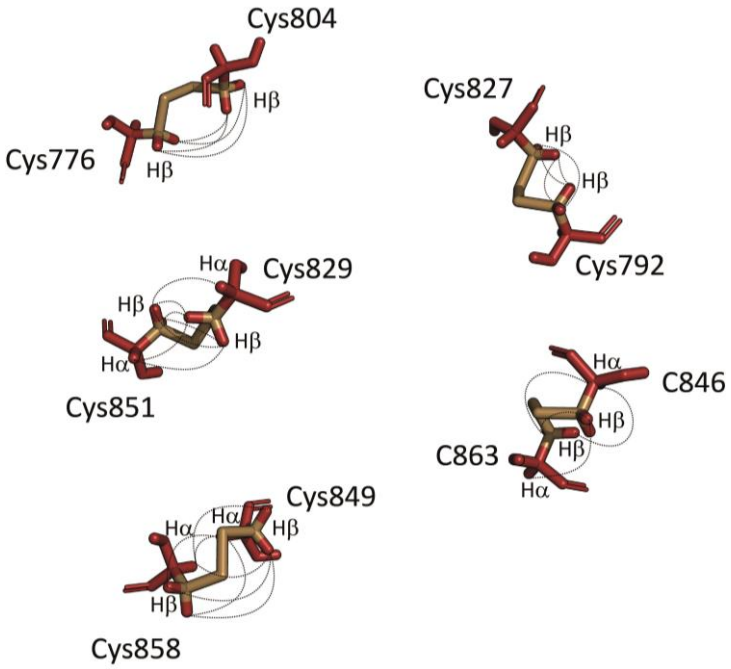


Figure 2.25: Across disulphide bond NOEs. Stick representation of the disulphide bonded TIL'E' cysteine pairs, showing across disulphide bond NOEs. The β carbons and sulphur atoms are coloured in gold. NOEs between Hα and Hβ atoms are depicted by dotted lines for Cys776-Cys804 (TIL'), Cys792-Cys827 (TIL'), Cys829-Cys851 (E'), Cys849-Cys858 (E') and Cys846-Cys863 (E').

2.10. RDC orientational restraints

In complement to NOEs, 361 vector-orientation restraints were also collected by means of residual dipolar couplings (RDCs). Orientation-dependent magnetic interactions (dipolar couplings) between atomic nuclei in isotropic solution conditions average to exactly zero as a result of Brownian rotational diffusion, where the inter-nuclear vector samples all possible orientations with equal probability. However, under anisotropic conditions, when a very slight deviation from the random distribution of macromolecules in solution is imposed (such as in a liquid crystal), not all molecular orientations are equiprobable, the molecule will thus have a preferential

orientation. In this case, the dipolar coupling does not average to zero and a residual fraction of the coupling survives (Figure 2.26). RDC interactions are very sensitive measures of the orientation of the corresponding inter-nuclear vectors, and thereby offer highly precise structural information[359,360]. In contrast to the short-range, local distance restraints derived from NOEs, RDCs provide global distance-independent orientational restraints[361].

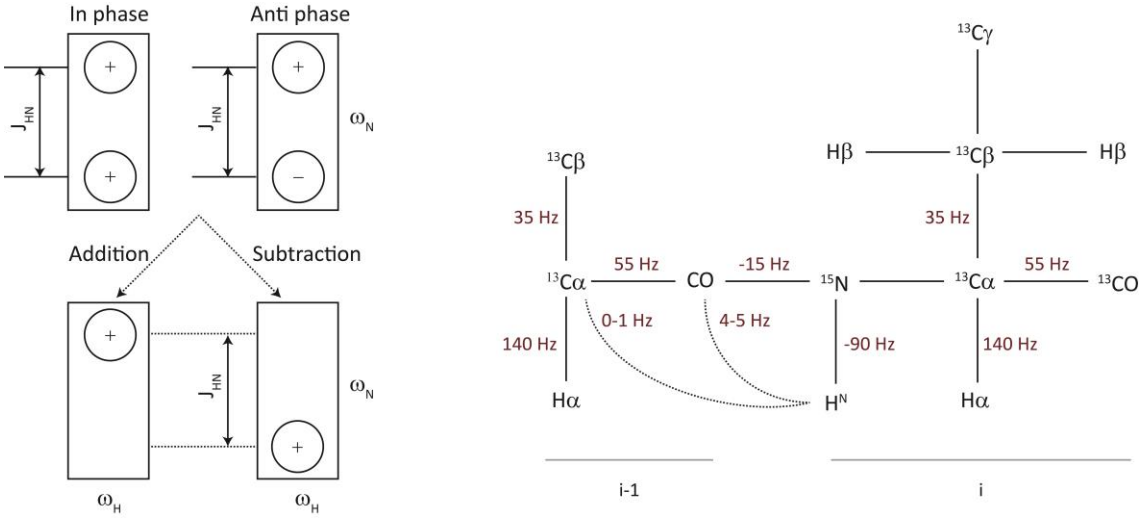


Figure 2.26: Extracting residual dipolar couplings. On the left is shown a schematic of a procedure for extracting $^1H^N_i$-$^{15}N_i$ couplings (J) under isotropic conditions from an IPAP (In phase/anti phase) ^{15}N-HSQC experiment, in which the dipolar couplings (D) average to zero. The ^{1}H-^{15}N coupling is allowed to evolve during the ^{15}N indirect chemical shift evolution period, leading to doublets for each backbone amide in the resulting spectrum. To prevent the increased spectral crowding due to doubling of peaks, two subspectra are recorded in which the splitting is either in phase or anti phase. Subsequent addition and subtraction of the subspectra gives rise to two further subspectra containing only the upfield or downfield component. The J coupling is then simply extracted by measuring the chemical shift difference between the upfield and downfield components for each amide. Subsequent repeat measurement under aligned conditions yields splittings that are the sum of the coupling constant (J) and the dipolar coupling (D), J+D. The RDC is then calculated by taking the difference between the isotropic coupling (J) and the partially aligned anisotropic splittings (J+D), to give the magnitude of the RDC (D). Beside the peptide backbone are shown the coupling constants (J) from which the effect of the residual dipolar coupling is calculated.

Although RDCs, impose tight restrictions on the orientation of the corresponding inter-nuclear vector, there results degeneracy in orientation, that is the inability to distinguish an isolated vector orientation from its inverse. Thus, multiple RDCs were measured within the peptide plane and in different alignment media to alleviate degeneracy[360,362]. Seven sets of RDCs were measured in a poly(ethyleneglycol)/hexanol mixture (that orientates the TIL'E' molecules by steric interactions); $^1H^N_i$-$^{15}N_i$, $^{15}N_i$-$^{13}CO_{i-}$

$_1$, $^1H^N_i$-$^{13}CO_{i-1}$, $^{13}CO_i$-$^{13}C\alpha_i$, $^1H^N_i$-$^{13}C\alpha_{i-1}$, in the TIL'E' backbone peptide plane; and $^{13}C\alpha_i$-$^1H\alpha_i$ and $^{13}C\alpha_i$-$^{13}C\beta_i$, at the junction of the peptide plane. $^1H^N_i$-$^{15}N_i$ RDC measurements were also repeated in filamentous phage, *Pf*1, that induces alignment of TIL'E' molecules by electrostatic interactions. It is of interest to note that *Pf*1 phage has a negatively charged surface and preparation of a weakly *Pf*1 aligned TIL'E' sample proved difficult due to a strong interaction of TIL'E' with the bacteriophage molecules; suggesting the presence of a positively charged region on TIL'E'.

RDC restraints were used to refine the orientations of the individual bond vectors, this improved both the β-sheet arrangement and precision of the overall TIL'E' structure. The orientational dependence of RDCs was also exploited to determine the relative orientation of domain TIL' with respect to E', as knowledge of the geometry of subunit assembly is critical to understanding the interface formed by TIL'E' in VWF·FVIII complex formation. Amino acid residues undergoing significant internal dynamics (as distinguished by their heteronuclear ^{1}[H]-^{15}N-NOE value; section 2.13) were excluded from the analysis. The quality of the RDC-refined TIL'E' structure was evaluated by comparison of back calculated RDCs from the lowest energy model against the experimentally derived RDC data, using the programme Module 1.0[363]. The resulting fits are of good quality for the greater proportion of RDCs measured; the high correlation is also graphically visible in Figure 2.27. Of note, where an increased scatter is observed in the plots (Figure 2.27), this is reflective of an increased relative error in the experimentally derived RDCs, due to the decreased magnitude of the measured RDC. The plots indicate that all RDC restraints are very close to being fulfilled and the RDC-refined TIL'E' model is an excellent representation of the molecule in solution. The correlation between the back-calculated RDCs and experimentally derived $^1H^N_i$-$^{13}C\alpha_{i-1}$ RDCs, however, is very poor. This can be reasoned to the very small dipolar couplings of only 0-1 Hz, from which it is difficult to derive an accurate RDC measurement (Figure 2.26), and is not the result of an incorrect TIL'E' model, as a single template was used for the back-calculation of all RDC measurements (Figure 2.27).

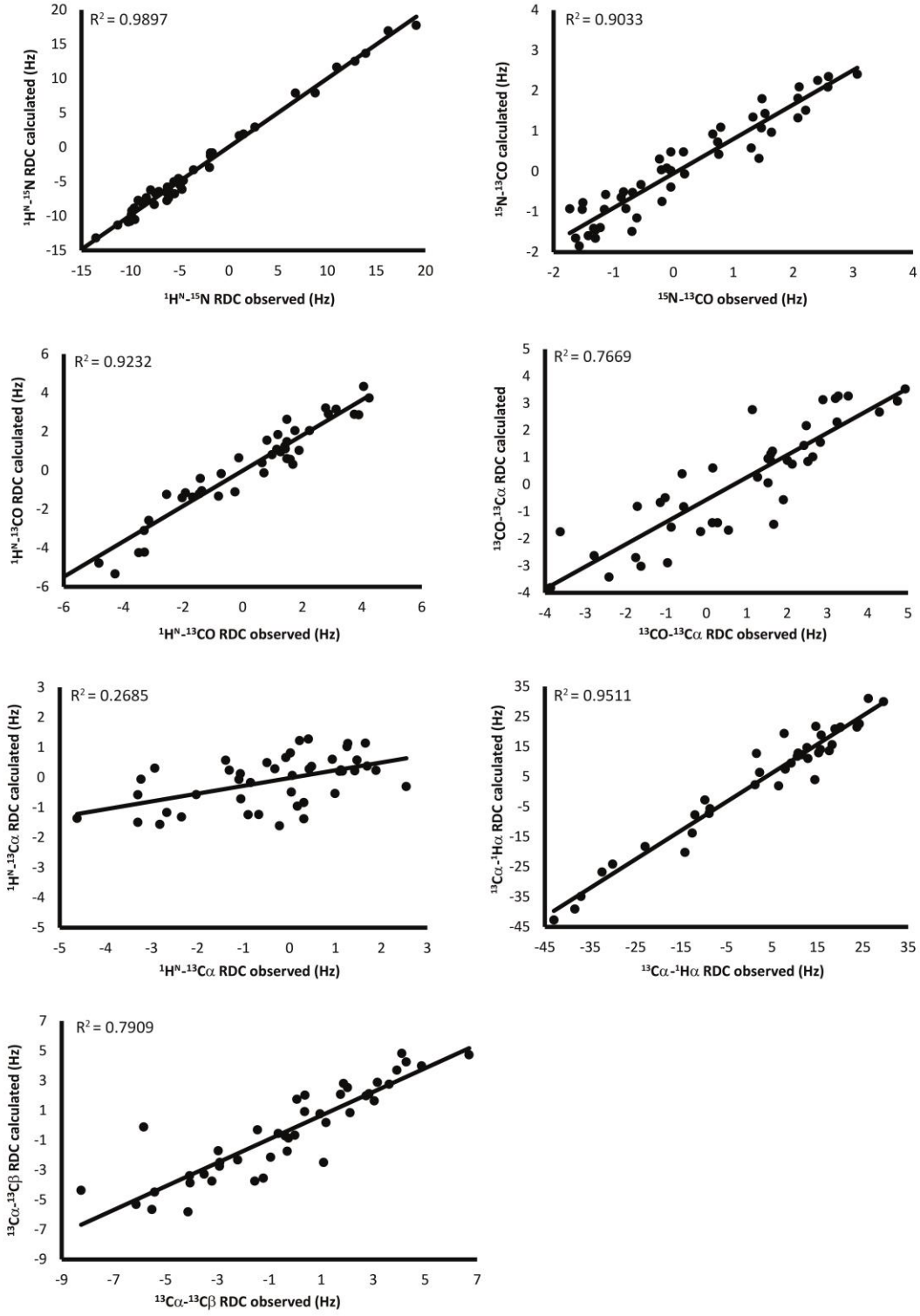


Figure 2.27: Correlation between experimental RDCs and back-calculated RDCs. Shown are correlation plots of the experimentally derived RDCs against those obtained from the RDC-refined final solution structure of TIL'E'. A single template TIL'E' structure was used for all calculations. The agreement is quite good, for most RDCs measured, with the exception of $^{1}H^{N}$-$^{13}C\alpha$ RDCs. The squared of the correlation coefficient (R^2), a measure of the goodness of fit of the trend line, is shown in the corner of each plot. Data for *Pf*1 aligned TIL'E' is not shown.

2.11. High resolution structure of VWF TIL'E'

The final high resolution solution structure of TIL'E' was calculated based upon the following experimental NMR restraints; 954 NOEs, 361 RDCs, 108 backbone dihedral angles and eight disulphide bonds, using the programme CYANA 3.0[364]. From the final set of calculations a family of the ten lowest energy structures was chosen to represent the solution conformation of VWF TIL'E'. The lowest energy solution structure of VWF TIL'E' (D' region) is shown in Figure 2.28.

The global fold of TIL'E' is crescent-shaped; this is consistent with the overall domain shape seen in low resolution electron microscopy imaging[140]. The structure clearly shows that TIL' and E' form independently folding domains and the domain assignment separating D' into two distinct domains is demonstrably correct.

Salient features of domain TIL' (aa 767-827) are the restricted regions of secondary structure (~ 84% loop, 16% β-strand) and lack of a notable hydrophobic core. Five conserved disulphide bonds constitute the major stabilising force constraining the fold, as is typical for small disulphide-enhanced domains[321,329]. The heart-shaped TIL' (length ~35 Å, width ~20 Å) is formed of two almost perpendicular β-sheets, β1:β2 (aa 772-775:806-809) and β3:β4 (aa 814-817:820-823) that form a scaffold bearing a long 30 residue loop, the β1-to-β2 loop (Figure 2.28). This dominating loop (aa 776-805) connects strands β1 and β2 and encompasses an eight residue insertion between the second and third conserved TIL family cysteines (Figure 2.2), that creates an apex in the VWF TIL' structure, the apical β1-to-β2 loop (Figure 2.28).

The second antiparallel β-sheet, formed by strands β3 and β4, are connected by a reverse turn forming a small hairpin structure. There is a single turn of a 3_{10}-helix (aa 824-826), with characteristic $d\alpha H^N(i, i+2)$ NOEs, at the extreme N-terminus of TIL' that carries forth into domain E'. Solvent accessible surface calculations, determined by the Ooi number[365], fail to identify a hydrophobic core and the TIL' disulphide bonds act as architectural cross-braces reinforcing and stabilising the tertiary structure. Bridge $Cys_1$767-$Cys_7$808, orientates the N-terminus with respect to strand β2 and imposes constraints on its conformation, bridges $Cys_2$776-$Cys_6$804 and $Cys_3$788-$Cys_5$799 stabilise the β1-to-β2 loop that is anchored to the TIL' C-terminus by bridge $Cys_4$792-

$Cys_{10}827$, and lastly, bridge Cys_8810-Cys_9821 serves to orient the two antiparallel β-sheets with respect to one another (Figure 2.28).

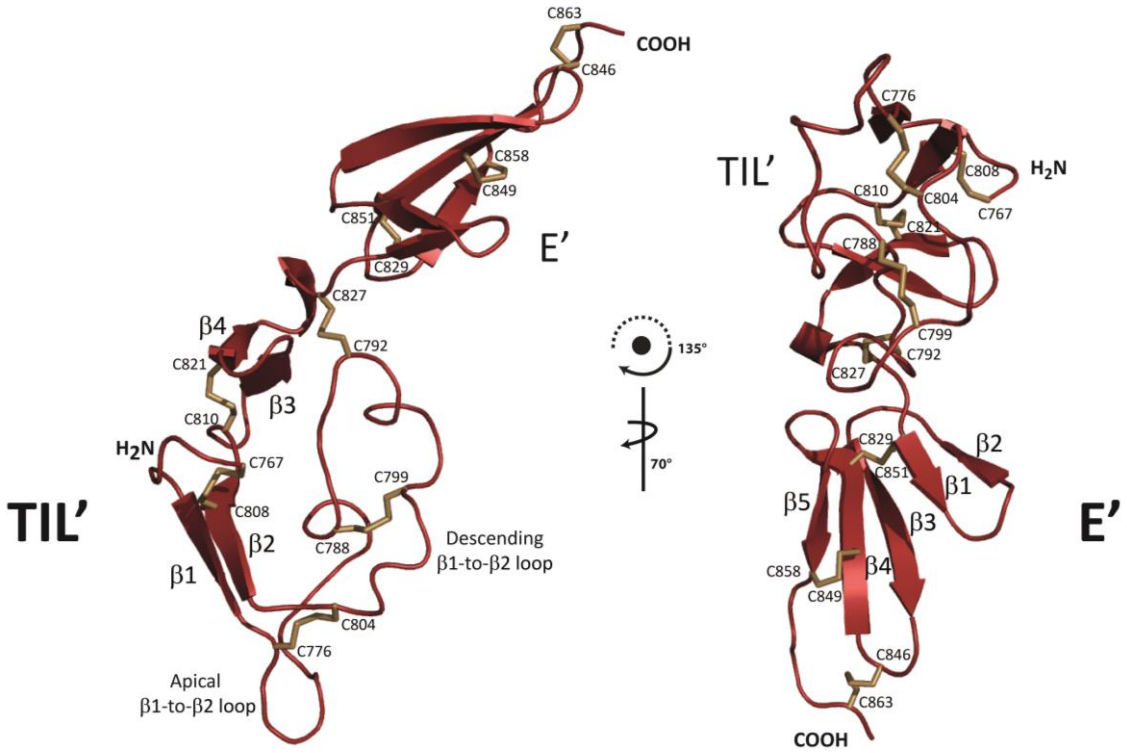


Figure 2.28: High resolution solution structure of VWF TIL'E'. A ribbon diagram of the lowest energy solution structure of TIL'E' is shown. Two orientations are presented for clarity of appearance of the individual domains; TIL' is dominant on the left and E' is dominant on the right. Stick representations of disulphide bonds are coloured gold and their respective cysteines labelled. β-sheets are numbered in the order in which they occur in the individual domains. Regions of the long TIL' β1-to-β2 loop are designated descending or apical for ease of orientation.

In well-defined contrast, the greater proportion of domain E' (aa 829-863) is formed of a triple-stranded all anti-parallel β-sheet; strands β3, β4 and β5 (length, ~25 Å; width, ~10 Å). The disulphide bond connectivity was unambiguously assigned as Cys_1829-Cys_4851, Cys_2846-Cys_6863 and Cys_3849-Cys_5858. The N-terminus of E' contains a short double stranded β-sheet, formed of strands β1 (aa 829-831) and β2 (aa 834-836), that is orientated with respect to strand β3 (aa 839-844) by the Cys829-Cys851 disulphide bond. Bridge Cys_3849-Cys_5858 retains strands β4 (aa 847-852) and β5 (aa 855-858),

which are collaterally supported by $Cys_1$829-$Cys_4$851 and $Cys_2$846-$Cys_6$863 (Figure 2.28).

The difference between the molecular surface area of the individual TIL' and E' domains (6,601 $Å^2$ and 3,977 $Å^2$, respectively) and that of the TIL'E' molecule (10,879 $Å^2$) is only 301 $Å^2$, indicating that there is minimal interaction between the two domains; this is further supported by the lack of inter-domain NOEs. The extension of domain E', away from TIL', is likely due to the one residue linker, Pro828, between the Cys boundaries of TIL' and E'. This proline, together with the lack of a significant linking loop, may impart rigidity that maintains the relative orientation of the two domains.

Although molecule TIL'E' lacks a definitive hydrophobic core and the residues are predominantly solvent exposed, there are small clusters of largely buried residues in domain TIL', namely at residues Thr789-Cys792, Gly807-Cys810, Met814-His817 in TIL' and Cys827-Cys829 at the TIL'-E' border (Figure 2.29).

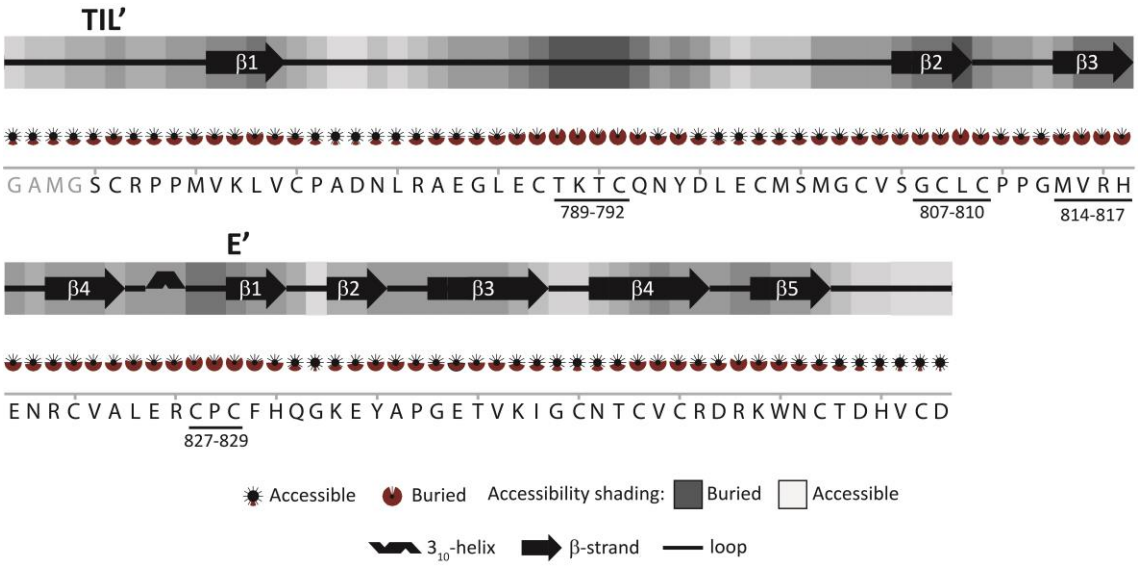


Figure 2.29: Accessible molecular surface of TIL'E'. Shown is the average estimated accessibility of the TIL'E' family of ten lowest energy structures generated with the programme Procheck-NMR[367] (implementing the Ooi number method). Accessibility is calculated from the the Ooi number, which is simply the number of Cα atoms within a 14 Å radius[365].

Interestingly, these residues appear to form a concave cleft-shaped pocket on the front face of the TIL' molecule. This pocket is bound laterally by strand β2 on the left and an ascending segment of the β1-to-β2 loop on the right; the superior border of the

pocket is formed by strand β4 and the upper portion of loop β1-to-β2 (Figure 2.30, left panel). Protein surfaces contain clefts and indentations of varying sizes and depths, and it is often the case that binding sites are found within the largest cleft on the protein surface[366]. Therefore, the buried pocket identified on the front face of TIL' is likely to be of functional significance (Figure 2.30, right panel).

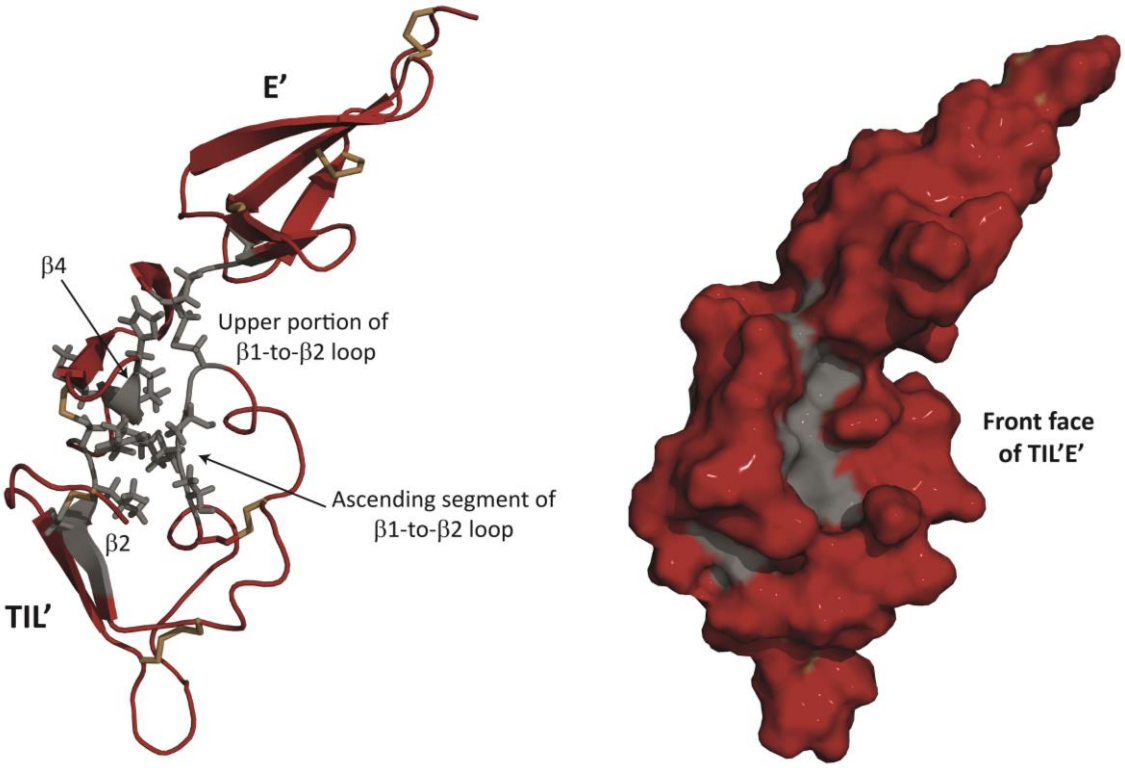


Figure 2.30: A buried pocket on the TIL' surface. Residue clusters identified by accessibility analysis to be buried are: Thr789-Cys792, Gly807-Cys810, Met814-His817 in TIL' and Cys827-Cys829 at the TIL'-E' border. On the left is shown a ribbon representation of structure TIL'E'. Buried residues are mapped into the TIL'E' structure and their side chains are represented by sticks (coloured grey). The buried residues form a pocket that is bordered laterally by strand β2 on the right and an ascending segment of the β1-to-β2 loop on the left, superiorly the pocket is bound by strand β4 and the upper portion of the β1-to-β2 loop. A surface representation of structure TIL'E' is shown on the left, where it is clear that these resides form a concave cleft-shaped pocket on the front face of TIL' (coloured grey).

The family of ten lowest energy TIL'E' structures, with the exception of the extreme N- and C-termini, is well-defined, and has an average backbone pairwise RMSD (root mean square deviation) of 1.12 Å; the corresponding mean RMSDs of the TIL' and E'

domains in isolation are 1.37 Å and 0.34 Å, respectively. The RMSD of the TIL' domain excluding the β1-to-β2 loop is 0.81 Å. These differences are distinctly evident when the TIL'E' backbone RMSD of each residue is plotted (Figure 2.31). Large RMSD deviations are strictly confined to domain TIL', with a prominence in the TIL' β1-to-β2 loop (Figure 2.31). Thereby, indicating that the β1-to-β2 loop (aa 776-805) dominates the structural heterogeneity of TIL' and alludes to its intrinsic dynamic character (Figure 2.32).

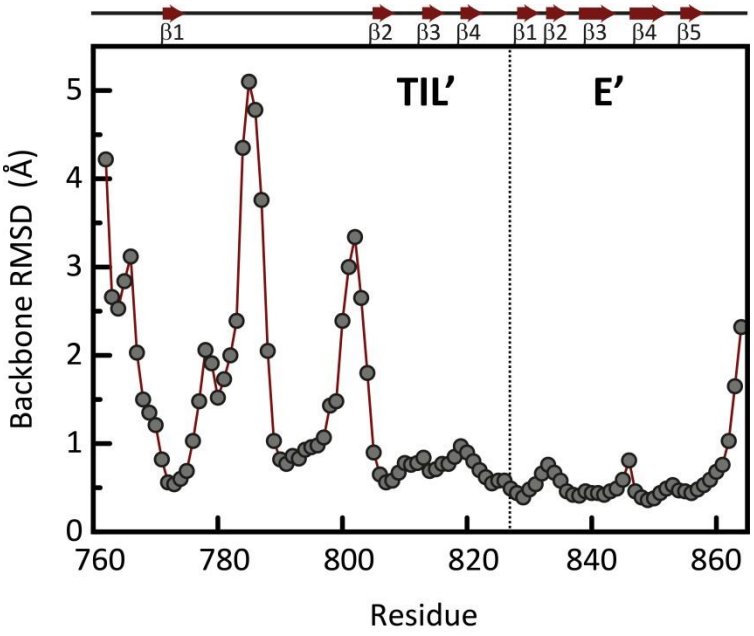


Figure 2.31: Per residue Backbone RMSD of the ten lowest energy TIL'E' structures. The backbone RMSD calculated on a per residue basis for the family of ten structures with the lowest energy. Strikingly evident is that residues of the β1-to-β2 loop of domain TIL' (aa 776-805) are significantly less well defined that the rest of the protein. RMSD plotted by Dr D. Flemming Hansen.

Analysis of the geometric quality of the TIL'E' ten lowest energy family reveals that it is composed of high quality structures (Figure 2.32), from which 95.4% of residues lie in the most favoured and additionally allowed regions of the Ramachandran plot[367] (73.1% in the most favoured and 22.3% in the additionally allowed region), and only 3.4% and 1.2% in the generously allowed and disallowed regions, respectively (Appendix 5.3). Small disulphide-rich protein folds typically possess less secondary structure than larger proteins, and therefore the percentage of residues found in the most favoured regions of the Ramachandran plot for this type of structure is usually lower relative to high resolution crystal structures of larger proteins[353]. Nevertheless, low backbone RMSD between the family of ten lowest energy structures and their

Ramachandran statistics confirm that the calculated TIL'E' structures are a precise representation of the experimental data, and hence, of the TIL'E' molecule in solution.

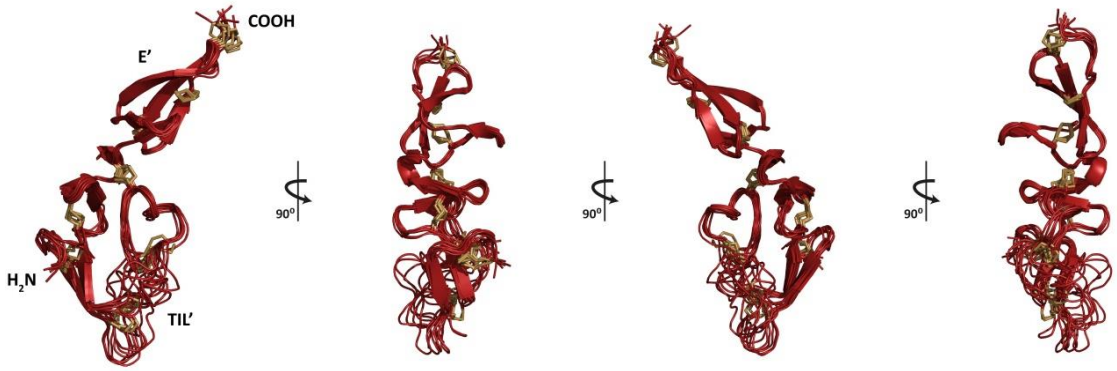


Figure 2.32: The family of ten lowest energy TIL'E' structures. Ribbon representation of the TIL'E' family of ten lowest energy structures are shown in four rotations of 90° about the y-axis. It is clear that the TIL' β1-to-β2 loop is conformationally heterogeneous whereas the β-sheet regions of TIL' and E' as well as the top of the TIL' β1-to-β2 loop possess a high degree of definition.

2.12. Structural homologues of TIL'E'

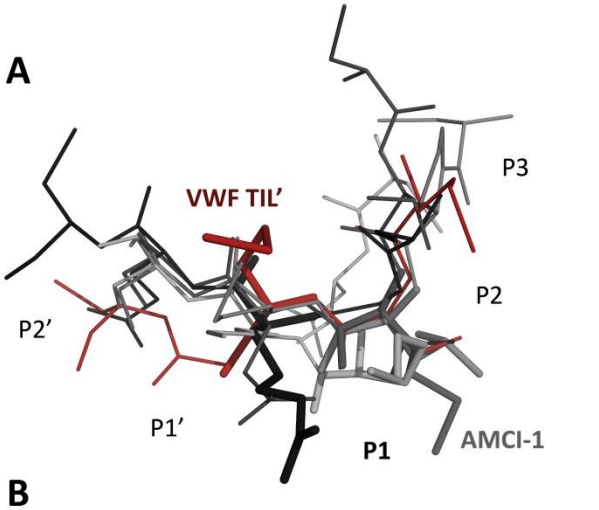
The VWF TIL' domain is a structural homologue of the *Ascaris* protease inhibitor (TIL units) family. Serine protease inhibition by the TIL family is accomplished by the Laskowski mechanism, binding to the enzyme in a substrate-like fashion. TIL domain reactive site residues lie between the fifth and sixth conserved TIL family cysteines[108], and are part of an exposed binding loop connecting strands β1 to β2, surrounding the hyper exposed P1 residue[112]. Stabilisation of the binding loop conformation, between Cys_5 (P3) and Cys_6 (P2'), is facilitated by two flanking disulphide bonds, linkages 3-5 and 2-6. Disulphide loops of cysteines 1-7, 2-6 and 4-10 encompass the scissile bond during conversion from virgin to modified inhibitor, to ensure that the two peptide chains do not dissociate[111].

In the VWF TIL' domain, position P1 corresponds to residue Met802 (Figure 2.33, panel A) and is solvent exposed, as is essential for hypothetical penetration of the P1 side chain into the S1 specificity pocket of a protease during complex formation. The dominant force for TIL inhibitor P1 binding is hydrophobic interactions, therefore excess steric bulk, β-branching (Val and Ile), polarity and charge at this position are

deleterious to complex formation[112,113,368]. This leaves only TIL', of the five VWF TIL domains, without physiochemical restrictions on its inhibitory potential, and a P1 Met is generally thought to inhibit chymotrypsin and chymotrypsin-like enzymes[111] (Figure 2.33, panel B). In addition, the reactive site loop is usually dynamic, as is evident for the VWF TIL' β1-to-β2 loop, in the uncomplexed state and becomes significantly rigidified upon complex formation[369].

Discussion of the potential inhibitory activity of domain TIL', however, is made with extreme caution as reactive-site residues in inhibitor units are often hyper-variable, thus, it is seldom possible to tell from the sequence whether or not an inhibitor homologue is likely to have inhibitory activity[108]. Furthermore, the VWF TIL' domain contains an eight residue insertion in the β1-to-β2 loop, between the second and third conserved cysteines, that is not present in other members of the *Ascaris* family (Figure 2.2), and the side chain of the P1 position of TIL' (Met802) is directed towards the core of the domain in half of the ten lowest energy TIL'E' conformers. Unlike the P1 conformation of other *Ascaris* protease inhibitors (Figure 2.33, panel A), TIL' Met802, in the lowest energy structure, does not significantly protrude from the protein scaffold and is partially buried, making it is difficult to conceive that in this position, Met802 is capable of serving as a protease recognition motif (Figure 2.33, panel C). Lastly, the function of VWF has been extensively studied, and as yet a serine protease inhibitory activity has not been found.

Experimental studies are required to resolve the question of VWF TIL' putative serine protease inhibition, as interestingly, VWF is perfectly placed for interaction with serine proteases of the coagulation cascade, of which factors XIa, Xa, IXa and VIIa, part of the S1 family that are inhibited by *Ascaris*[107].



B

	pdb	P3	P2	P1	P1'	P2'
AMCI-1	1CCV	Cys	Thr	Met	Gln	Cys
BSTI	1HX2	Cys	Asp	Lys	Lys	Cys
ATI	1ATA	Cys	Thr	Arg	Glu	Cys
C/E-1	1EAI	Cys	Pro	Leu	Met	Cys
NAPc2	1COU	Leu	Val	Arg	Val	Cys
NAPc5	2P3F	Arg	Ser	Arg	Gly	Cys
VWF TIL'	-	Met	Ser	Met	Gly	Cys

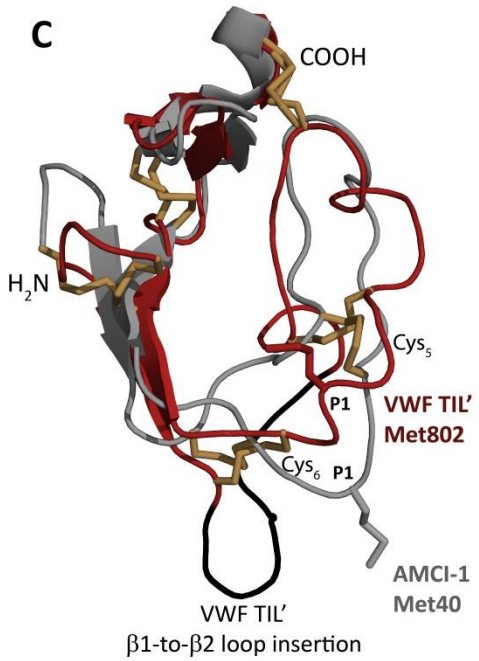


Figure 2.33: Reactive site residues of the *Ascaris* trypsin inhibitor-like family. Panel A shows the superposition of VWF TIL' positions P3 (Met800) to P2' (Cys804) coloured red, with the corresponding reactive site residues of AMCI-1[317], BSTI[319], ATI[318]and C/E-1[316], that inhibit either chymotrypsin or trypsin, coloured in increasingly darker shades of grey. These are shown in stick representation, with the P1 position shown with a larger stick radius for emphasis. Hydrogen atoms are removed for clarity. A comparison of the binding site sequence in the β1-to-β2 loop of all *Ascaris* inhibitors, for which a high resolution structure is known, are shown in panel B. In panel C is a superposition of the VWF TIL' NMR structure (red) with that of AMCI-1 (grey; pdb: 1CCV[317]). The P1 positions, bound by conserved Cys_5 and Cys_6, are indicated and shown as stick representations; VWF TIL', Met802 and AMCI-1, Met40. The eight residue insertion on VWF TIL' between conserved Cys_2 (Cys776) and Cys_3 (Cys788) is coloured black on the VWF TIL' structure.

The VWF E' domain adopts the same fold as subdomain 1 of VWC modules, belonging to the fibronectin I-like superfamily of disulphide-rich, all β-structures that have arisen from a VWC ancestor. The prototypical VWC domain is an extended rod-like structure whose fold manifests a two subdomain architecture, termed subdomain 1 and subdomain 2[101,118]. N-terminal subdomain 1 is well-structured and of predominantly β-sheet configuration, from which the E' domain evolved as a truncation. C-terminal subdomain 2 however, does not contain elements of secondary structure and the loop is constrained by disulphide cross links. Structure of the complex between the VWC module of crossveinless-2 and its cognate ligand (BMP-2), indicates that subdomain 1 forms the functional subdomain with disulphide linkage 3-5 at the centre of the

binding interface[101], however, recent data suggests an additional and distinct binding role for subdomain 2 of this VWC module[370], thereby highlighting the versatility of VWC domains, and the spatial variability in binding epitopes.

Superposition of N-terminal VWC subdomain 1 from human procollagen IIA[118] and VWF domain E' shows a clear fold level similarity with an RMSD of 1.33 Å (pairwise sequence identity, 30.6%; similarity, 38.9%) (Figure 2.34).

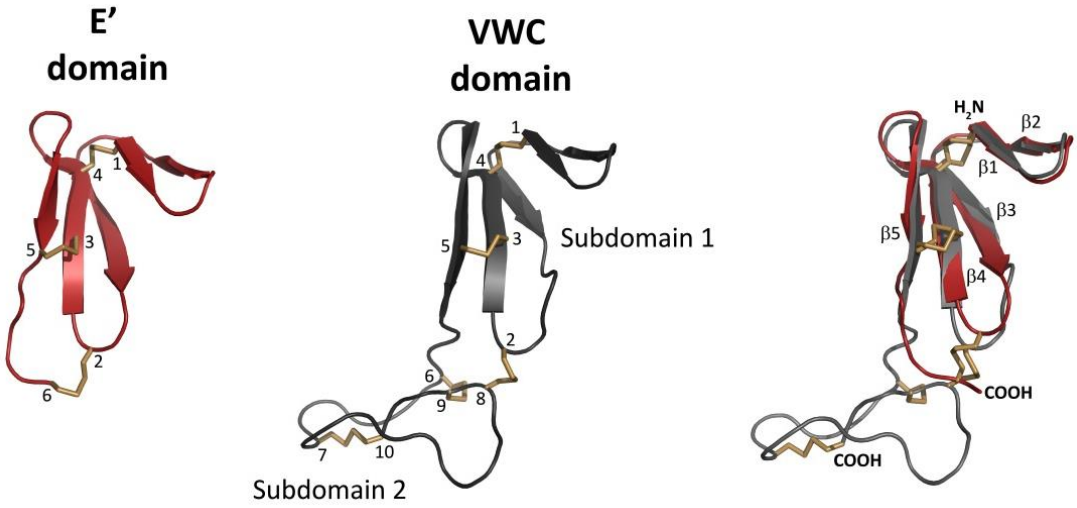


Figure 2.34: VWF domain E' and subdomain 1 of VWC domains adopt the same fold. The three-dimensional lowest energy structure of VWF E' (far left; red) and the NMR structure (pdb: 1U5M)[118] of von Willebrand type C (VWC) domain of human procollagen IIA[118] (centre; grey), together with their superposition (far right) are shown in ribbon representation. Disulphide bonds are coloured gold and conserved Cys residues are numbered according to the order they occur in the protein sequence.

Strong structural similarity is highly indicative of a divergent evolutionary relationship[124], however, what this structural ancestry implies for protein fold and function space is not clear[124,371]. Whilst the coincidence of the same major structural elements in the same arrangement and topological connections implies functional inheritance[372], reliable predictions still cannot be made.

To retain a specific function through evolution, the local environment of a functional site must be preserved, even if other portions of the fold have been altered[372]. However, in view of the inter-cysteine sequence divergence of VWC domains and hence, inability to predict binding epitopes based on sequence alone[118], coupled with the paucity of data regarding VWF C domain functions, the possibility of deriving a

sensible hypotheses for the function of VWF E domains, becomes increasingly difficult. For most VWC modules, the cellular role has yet to be determined, however, regulating BMPs and TGF-β signalling are the two most common functions attributed to VWC modules. There is emerging evidence that lateral self-association of VWF, forming large fibrous networks at sites of vascular injury, may confer an enhanced capacity for platelet capture by immobilised VWF, and this is mediated by VWC3 subdomain 1 Cys2431-2453 and Cys2451-Cys2468 allosteric disulphide bonds (linkages 1-4 and 3-5)[178,358,373]. In addition, under simulated *in silico* stretching, fibronectin I domains exhibited significant elongation and internal structural rearrangements that include changes in disulphide bond configuration. The extension of disulphide bonds under tension, is thought to contribute to the overall elongation of fibronectin I polymers in the extracellular matrix[374]. Structurally related clusters of proteins may be characterised by a specific distribution of functions unique to that fold[375], and for that reason, the VWC subdomain 1 and E domain disulphide bonds may be revealed to facilitate shear dependent VWF extension, in which alignment of disulphide bonds along the axis of flow lead to local domain level extensions, the summation of which, unravel the globular VWF configuration.

In summary, the high resolution solution structure of TIL'E' was solved, and the disulphide bond topology unambiguously assigned. This represents the fundamental basis upon which a model of VWF·FVIII complex formation may be built, and the structural and functional consequence of type 2N VWD mutations may be inferred.

2.13. Conformational dynamism of TIL'E'

Protein dynamics is intimately linked to molecular function, hence, the dynamic characterisation of VWF TIL'E' is of great importance for elucidating the mechanism by which FVIII is stabilised in plasma. The most severe approximation to structure determination is the assumption that a protein can be represented by a single structure. It has long been recognised that stably folded proteins are dynamic, displaying motions across a broad hierarchy of timescales[376–378] (Figure 2.35). Biological processes are often highly dependent upon and governed by,

conformational transitions, and there exists an intimate relation between protein motion and function. NMR is the most powerful technique for deconvoluting the dynamic motions of proteins, thus providing a more accurate representation of the conformational space sampled by a biomolecule and permitting the reconciliation of structure, dynamics and function[376,379]. As the symbiosis between protein dynamics and structure is essential to biological activity[380], the structural dynamics of TIL'E' were quantified experimentally by ^{15}N nuclear spin relaxation experiments to provide mechanistic insight into TIL'E'·FVIII binding. Analysis of relaxation data reveal interesting motional trends in TIL'E'; most strikingly, that there is a readily distinguishable mobility differential between TIL' and E'.

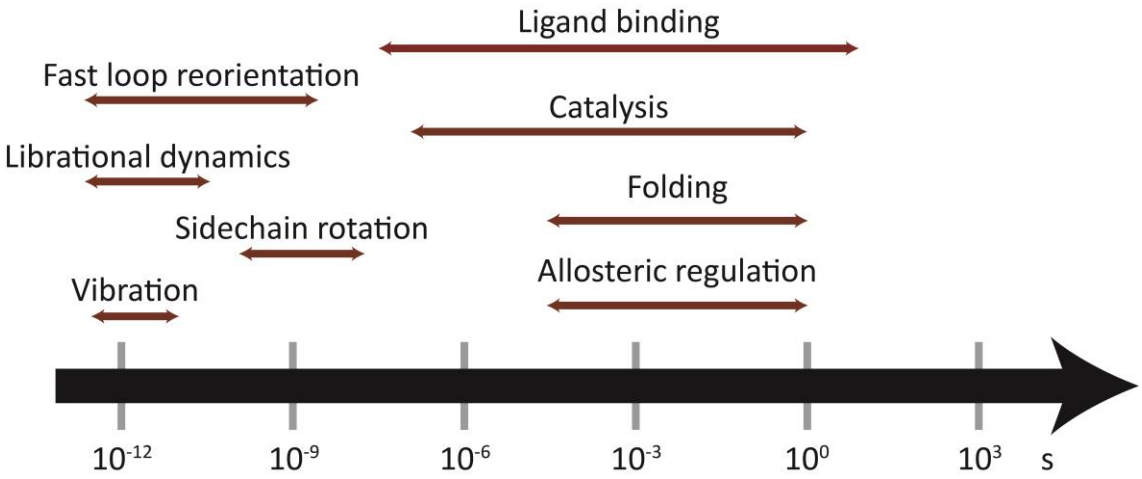


Figure 2.35: Time scales for protein dynamics. A schematic of the various molecular motions that can be probed by NMR experimentation, these span more than 12 orders of magnitude in timescale.

Backbone fluctuations on a sub-nanosecond timescale (picosecond to nanosecond) were measured by heteronuclear [^{1}H]-^{15}N NOEs (hetNOEs). Fast timescale motions act to stabilise the protein in its folded state and their presence is a necessary prerequisite for slower timescale dynamics involving large-scale collective motions[376].

The hetNOE value determined for each TIL'E' residue is shown in Figure 2.36; hetNOE values below 0.6 indicate flexible, unstructured regions, whereas, values above 0.6 are consistent with rigid secondary structure. As for most proteins, the residues at the N- and C-termini of TIL'E' are disordered, with low hetNOE values. These data clearly demonstrate that domain E' is rigid, with a uniform distribution of hetNOE values around 0.65 (excluding the flexible C-terminal Asp864 residue) indicating low flexibility

and an ordered structure. Whereas, domain TIL' exhibits distinct localisation of residues in the TIL' β1-to-β2 loop with significantly reduced hetNOE values, namely residues Ala778, Asp779, Arg782, Glu784, Ser801 and Gly803, indicating a comparatively higher degree of flexibility in this region of TIL' (Figure 2.36). Furthermore, flexibility in the immediate vicinity of TIL' disulphide bridges is restricted, as evidenced by hetNOE values greater than 0.6 for all available Cys residues (Figure 2.36), thereby, confirming their importance as structural bridges that maintain the tertiary structure of the TIL' fold.

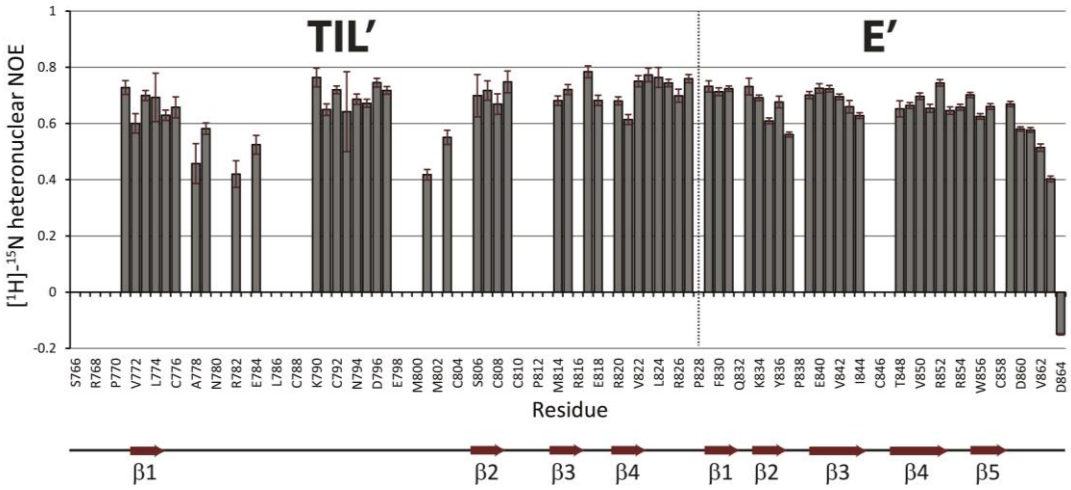


Figure 2.36: Backbone TIL' [^{1}H]-^{15}N NOEs: picosecond to nanosecond dynamics. Heteronuclear NOEs report on dynamics on the sub-nanosecond timescale. Heteronuclear amide NOE values are a measure for the dynamics of the local environment. Values below 0.6 indicate flexible unstructured regions, while stabilising structural elements give rise to values greater than 0.6. Secondary structure elements of TIL'E' are depicted below the plot for clarity. Spectrum obtained at 600 MHz at 25 °C. Data were not obtainable for all assigned residues due to weak or overlapped peaks.

VWF TIL'E' is evidently a dynamic structure; thus far however, the dynamics of TIL'E' have been isolated from the structure determination process. This gives rise to considerable uncertainty as to the distribution of conformations that the native TIL'E' molecule encompasses. Therefore, an ensemble of structures that sample all conformations satisfying the steric and dynamic NMR derived data was calculated. The TIL'E' conformational ensemble reveals distinct dynamicity in domain TIL'.

The RCI-derived S^2 order parameters, obtained from the backbone chemical shifts[345] (section 2.6, Figure 2.18), report on the amplitude of backbone motions on a timescale ranging from picoseconds to hundreds of microseconds[346]; RCI-S^2 values range from 0

(no orientational bias) to 1 (completely rigid)[379,380], thereby characterising the spatial restriction of the motion of each residue in TIL'E'. In addition, dynamics on the picosecond to microsecond timescale are also encoded in RDCs[381,382] which were combined with RCI-S^2 parameters, and subsequently included in structural calculations to determine an ensemble of protein conformations that represent, simultaneously, the TIL'E' structure and its associated dynamics (dynamic ensemble refinement). Accordingly, the structural ensemble represents a more faithful atomistic description of the conformational variability of the TIL'E' native state, on a picosecond to microsecond timescale[383] (Figure 2.37). Indeed, the TIL' β1-to-β2 loop is significantly more flexible than the remainder of TIL'E' and samples a heterogeneous ensemble of conformations that interconvert on the picosecond to microsecond timescale. These loop fluctuations facilitate larger-amplitude domain motions, as this region of TIL' was subsequently found to be coincident with slower dynamics, on the microsecond to millisecond timescale.

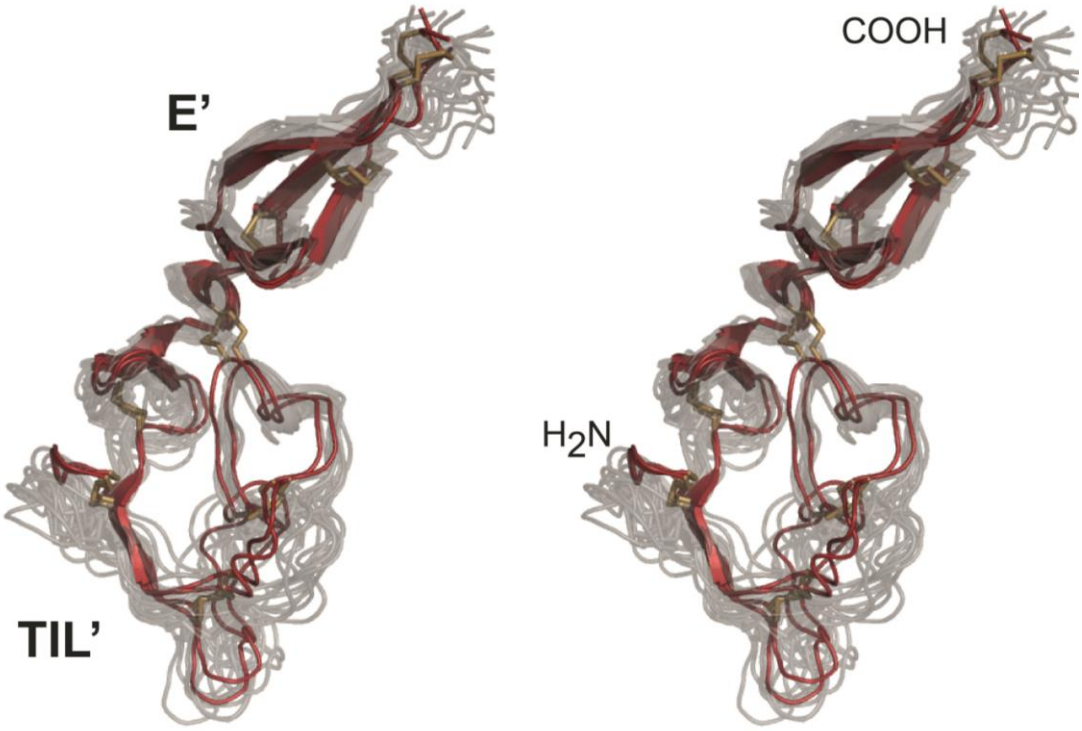


Figure 2.37: Stereo view of the structural variability in the TIL'E' ensemble. The method of dynamic ensemble refinement[383] was implemented to give an ensemble of TIL'E' structures that simultaneously represent the TIL'E' structure and its picosecond to microsecond dynamics. Disulphide bonds were added explicitly into the ensemble structure calculation. The twelve-member ensemble is shown in grey. This is overlaid with the two lowest energy structures from the single structure calculations (red). The view is of a cross-eye stereo. The ensemble refinement calculations were performed by Dr D. Flemming Hansen.

Conformational exchange on the millisecond timescale overlaps with the time constant of dynamic processes relevant to higher-order functions of macromolecules, such as ligand binding. Therefore, Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments, that specifically probe microsecond to millisecond dynamics and are sensitive to low-populated thermally excited states (conformers with much lower populations), were performed on TIL'E' to quantify such movements. The R_{ex} factor is derived from the CPMG relaxation dispersion experiments; R_{ex} is the relaxation rate due to conformational exchange, as nuclei sample various conformations and magnetic environments, and a value greater than 1.5 s^{-1} indicates microsecond to millisecond backbone mobility (conformational malleability). A first inspection of the CPMG data reveals dramatic compartmentalisation of slower motions in TIL', particularly in the β1-to-β2 loop (Figure 2.38). This confirms the existence of a TIL' conformational sub-state with a lifetime in the millisecond range, while the E' domain remains rigid on this timescale. Additionally, the presence of slower dynamics confirms that the incidence of a high RMSD in the ten lowest energy structures of TIL' is correlated, in part, with the motion of the molecule, rather than simply a lack of structural constraints (Figures 2.24 and 2.31).

A detailed analysis of the CPMG relaxation dispersions reveals that the ground state of TIL'E' interconverts stochastically with a sparsely populated excited state (conformer); the lifetime of the excited state is approximately 0.48 ms and it has a population of 3.5%. Because excited state conformers are only transiently formed and marginally populated, their structural characterisation eludes most biophysical methods. CPMG relaxation dispersion data, however, may be used to derive a structural model of the excited state and the dynamic processes observed experimentally[384].

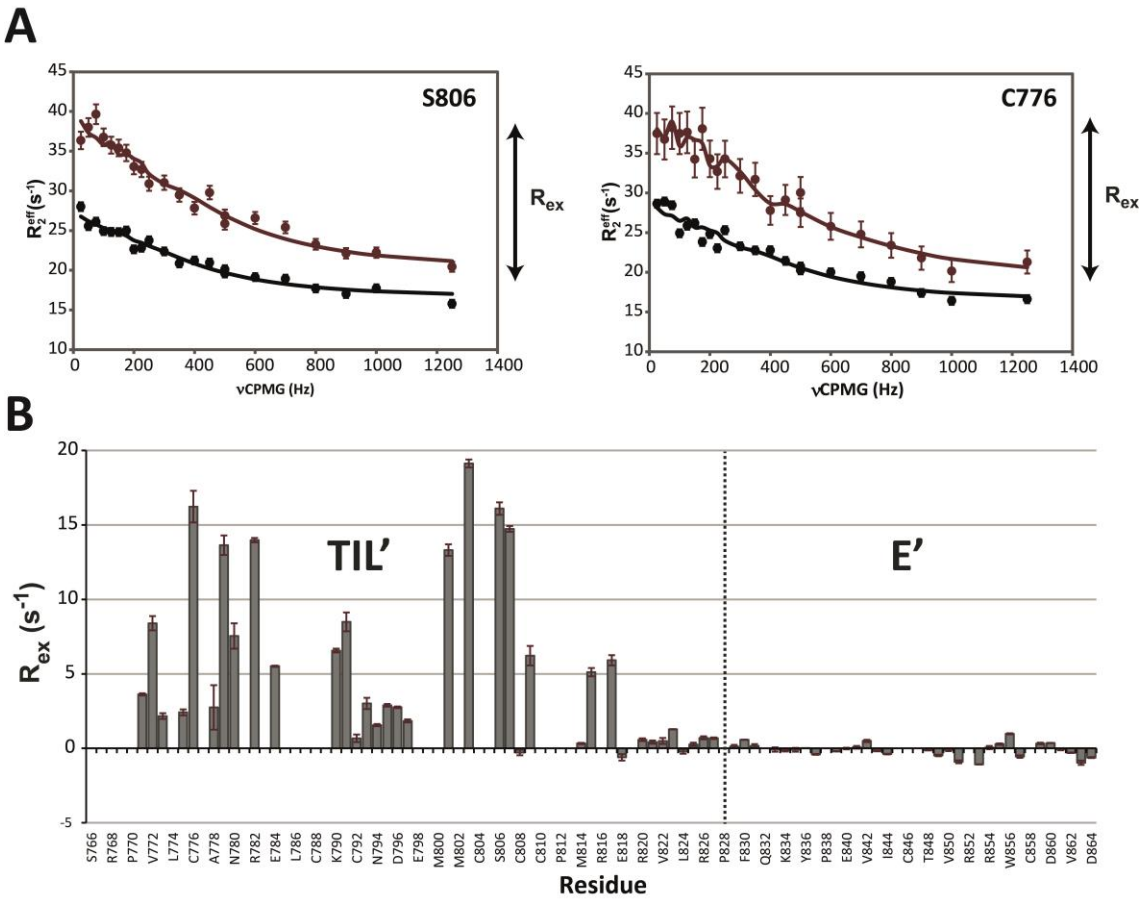


Figure 2.38: Microsecond to millisecond TIL'E' dynamics. Conformational exchange processes generally occur on the microsecond to millisecond timescale and thus report on slower motions than [^{1}H]-^{15}N heteronuclear NOEs or S^2 order parameters. CPMG experiments are made more robust by conducting the experiment at two external magnetic field strengths. Representative dispersion profiles are shown in the panel A for the conformationally dynamic residues Cys776 and Ser806 of TIL' (red, 700 MHz; black, 500 MHz). The chemical exchange contributions, R_{ex}, at 700 MHz are shown in panel B, where there is striking segregation of backbone millisecond dynamics in domain TIL', while E' is comparatively rigid on this timescale. R_{ex} values greater than 1.5 s^{-1} indicate dynamics on a microsecond to millisecond timescale (conformational exchange). Data were not obtainable for all assigned residues due to weak or overlapped peaks.

Conformational dynamism, as observed for VWF domain TIL', is an inherent property encoded into the protein fold through the evolutionary selection of conformers required for function[380]. Protein excited states often have significant functional roles in molecular recognition and ligand binding, such as in conformational selection of the excited state that is primed for binding to the cognate ligand, and is sought out from an ensemble of conformations that are already populated in solution[385]. The TIL'E' excited state may indeed be the conformer involved in recognition of the FVIII light chain in VWF·FVIII complex formation. Consistent with this theory is the finding that an

increase in temperature augments the VWF·FVIII associate rate constant[252], as it is thought that interactions between proteins with preoptimised binding sites benefit from an increase in the kinetic energy of the system, whereas, the association of proteins requiring conformational adaptations for formation of complex, in an induced-fit model, are hampered by increases in temperature[252,386]. Furthermore, it is interesting to note, that residues forming the lateral and superior borders of the buried pocket on the front face of TIL', identified by accessibility analysis (Figure 2.30), are dynamic on a microsecond to millisecond timescale (Lys790, Thr791, Gly807, Leu809, Val815 and His817) (Figure 2.38). Thus, conformational transition of the TIL' pocket, forming an optimised FVIII light chain binding site, supports a conformational selection model for VWF·FVIII complex formation.

Crucially however, the appropriate control is required to confirm the functional importance of the dynamics revealed in domain TIL', and to distinguish relevant dynamics from random fluctuations[387]. An approach would be to employ type 2N mutants and FVIII light chain binding experiments to alter the dynamics observed by NMR, so that changes in dynamics may be correlated with TIL'E'·FVIII complex formation.

A summary of the TIL'E' dynamics is shown in Figure 2.39 and detailed in Appendix 5.4, from which it is clearly evident that the TIL' β1-to-β2 loop is strikingly dynamic on at least two timescales, whereas, the remainder of TIL'E' provides a rigid scaffold supporting this conformationally dynamic entity. It is highly probable that the sequestration of dynamics localises the major FVIII binding site to the buried pocket on the front face of domain TIL'.

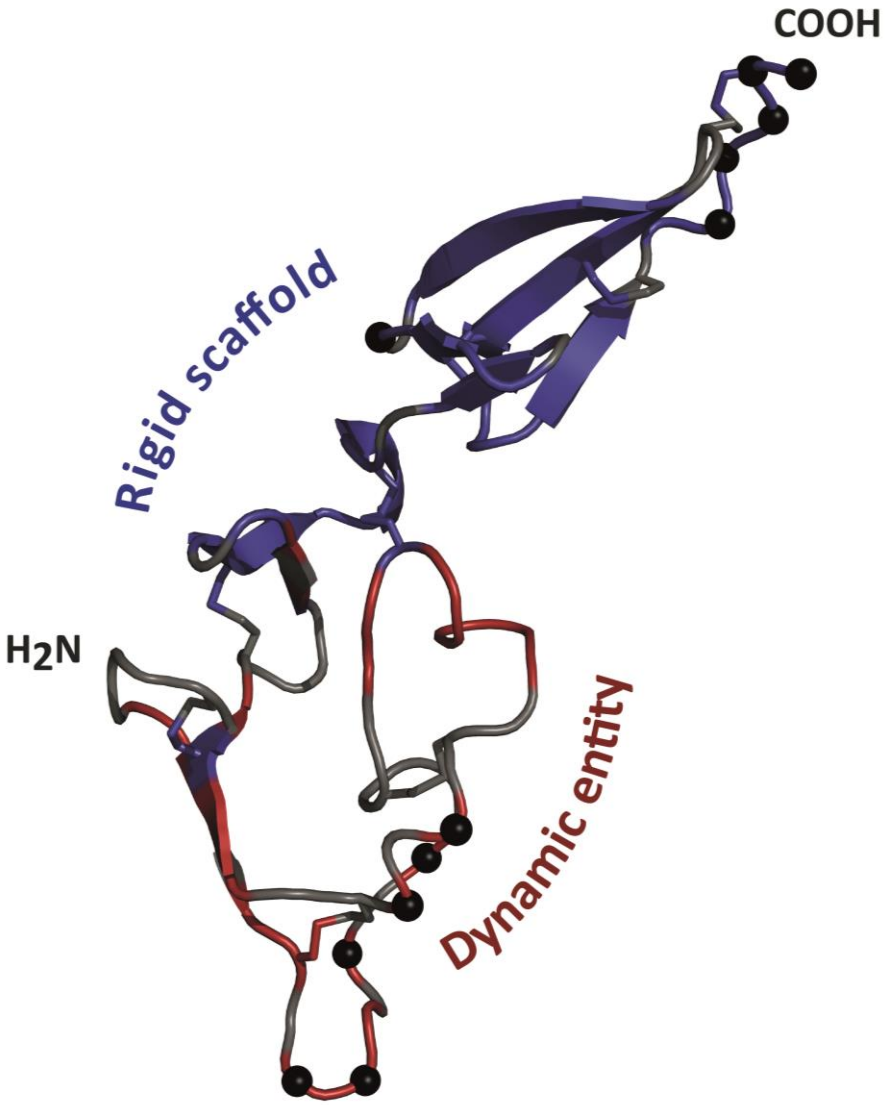


Figure 2.39: Summary of the dynamic character of TIL'E'. The lowest energy structure of TIL'E' is shown in ribbon representation. The combined dynamics of TIL'E', wherein black Cα spheres represent [^{1}H]-^{15}N heteronuclear NOEs with values < 0.6 (fast picosecond to nanosecond dynamics), and residues coloured red if R_{ex} is greater than 1.5 s^{-1} (slower microsecond to millisecond dynamics; conformational exchange) and blue is R_{ex} is less than 1.5 s^{-1}. Residues are coloured grey where no R_{ex} is available, either due to weak or overlapped peaks. It is clear that the β-sheets of TIL' and E' form a rigid scaffold supporting the highly dynamic TIL' β1-to-β2 loop.

2.14. Structural characterisation of type 2N VWD mutations

The solution structure and dynamic characterisation of VWF TIL'E' allows a first detailed structural interpretation of type 2N VWD missense mutations that disrupt the VWF·FVIII interaction; this provides the link between genetic pathology and clinical phenotype, and facilitates mapping of the FVIII binding region onto TIL'E'. The greater proportion of pathologic type 2N mutants are found in TIL'E', with distinct clustering in the TIL' β1-to-β2 loop, and are thus, coincident with the TIL' region of dynamicity (Figure 2.40).

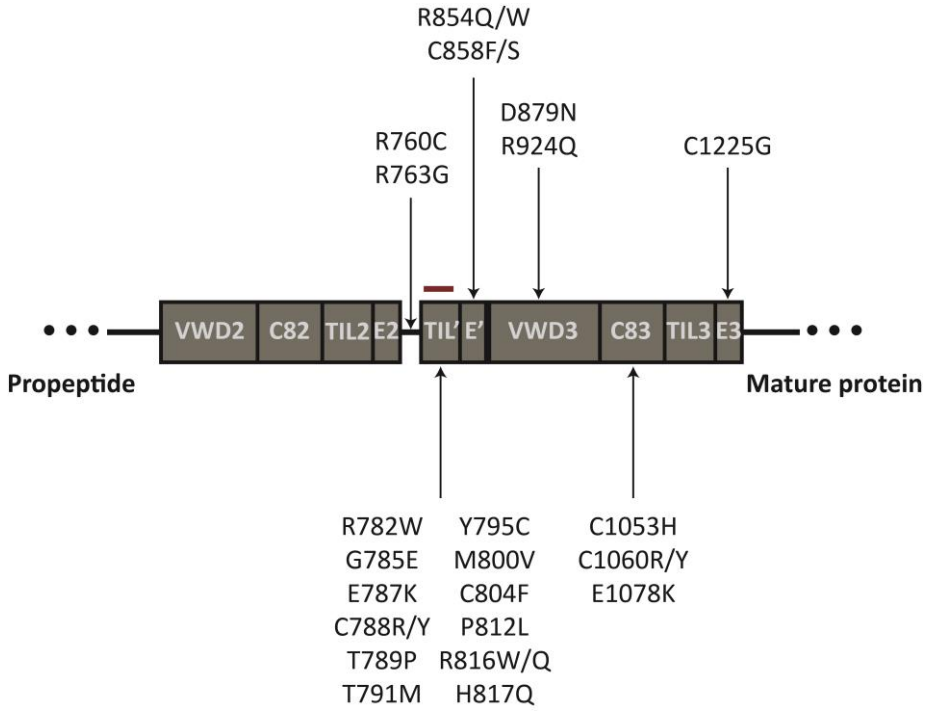


Figure 2.40: Type 2N VWD mutations. Schematic of the domain localisation of all reported type 2N mutations. Shown are prodomains VWD2-C82-TIL2-E2 (D2 region) and the N-terminal domains of mature VWF; TIL'E' (region D', exons 18-20) and VWD3-C83-TIL3-E3 (region D3, exons 20-28). Mutations are given with the one letter amino acid code. A red bar localises the epitopes of potent FVIII binding blocking monoclonal antibodies.

Interestingly, only two non-type 2N mutations are found in domain TIL'; type 1 VWD mutants Met771Ile and Arg782Gln. Of the remaining mutations reported[236], excluding the 18 type 2N mutations in TIL'E', two occur in the linker between the VWF prodomains and TIL'; Arg760Cys and Arg763Gly located at positions P4 and P1 of the consensus sequence of the furin site and obstruct propeptide cleavage and release of the mature VWF protein. In these cases of hereditary persistence of proVWF, it is

assumed that the attached propeptide sterically hinders FVIII binding to the adjacent TIL'E'VWD3 region[162,388,389]. Two type 2N mutations are in domain VWD3 (Asp879Asn and Arg924Gln), four in domain C83 (Gln1053His, Cys1060Arg/Tyr and Glu1078Lys) and lastly, one in domain E3 (Cys1225Gly). Interestingly, four of five mutations in domains C83 and E3, involve the substitution of a Cys residue (Figure 2.40).

Type 2N mutations in TIL'E' that involve a cysteine residue (Cys788Arg/Tyr, Tyr795Cys and Cys804Phe in TIL' and Cys858Ser/Phe in E') are associated with a triple biological defect, that of aberrant multimerisation (typical of type 2A VWD), secretion and FVIII binding[315,341,390,391]. Cys788, Tyr795 and Cys804 localise to the β1-to-β2 loop (Figure 2.41). Evidence that these mutations induce a deleterious conformational change has been obtained by means of antigen-capture experiments with conformation-dependent FVIII binding blocking mAbs[315,341], thereby substantiating the importance of the TIL' β1-to-β2 loop in FVIII binding and that correct disulfide bond formation is critical to the folding and function of VWF. Introduction of a Cys residue (Tyr795Cys) causes a deviant supranormal multimerisation pattern[315,391], whereas, breakage of disulphide bonds (Cys788Arg/Tyr, Cys804Phe and Cys858Ser/Phe) results in a reduction of high molecular weight forms[315,341,390]. Disulphide bond scrambling or suboptimal interaction with the VWF propeptide leading to aberrant multimerisation may block the secretion of higher molecular weight forms due to the multiplicity of errors with increasing molecular weight or there may be a dominant negative like effect, whereby heterozygous (and homozygous) dimers are retained in the ER; the reduction in the total number of dimers entering the Golgi compartments would lead to formation of only low and intermediate multimers, and consequently reduced secretion and FVIII binding, which are secondary effects.

Of the type 2N mutations (Figure 2.40), three account for the majority of cases reported – Thr791Met and Arg816Trp in domain TIL' and Arg854Gln in domain E' (Figure 2.41). Mutant Arg854Gln is the most common, in absolute terms, occurring with polymorphic frequency[10]. As with most of the qualitative VWD variants, type 2N demonstrates good genotype-phenotype correlations; Arg854Gln mutant is associated with a mild phenotype and FVIII levels of ~25%, while both the recurrent Thr791Met and Arg816Trp mutations result in FVIII levels of less than 10%[312]. All three of these

commonly reported type 2N substitutions are in immediate proximity to the TIL' β1-to-β2 loop (Figure 2.41). The Arg854 mutant in domain E', suggests an approximation of the E' β4-to-β5 loop to domain TIL', this has been confirmed by the TIL'E' solution structure (Figure 2.28 and Figure 2.41).

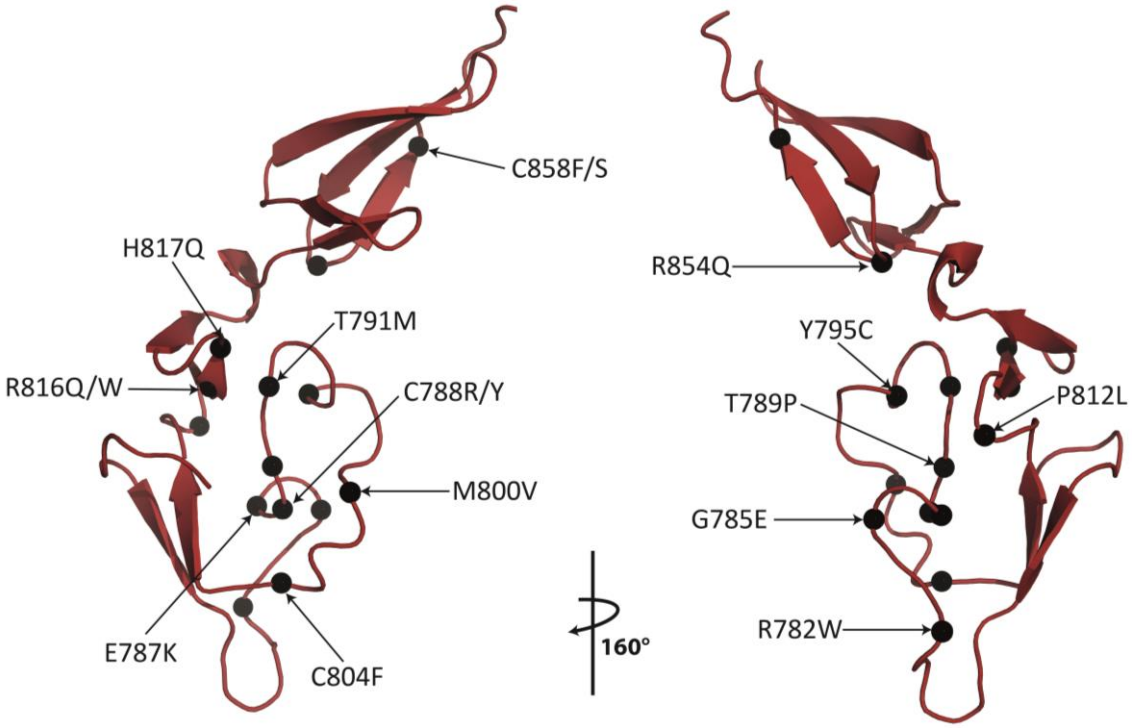


Figure 2.41: Type 2N VWD mutants in TIL'E'. A ribbon representation of the TIL'E' structure is shown, onto which all type 2N mutations reported in this region are mapped. Mutant residues are shown as black Cα spheres. Mutants are designated by their one letter amino acid codes. The structure on the right is rotated 160° about the y-axis, relative to the structure on the left. The structure on the left shows the front face of TIL'. There is clustering of type 2N mutants in the TIL' β1-to-β2 loop.

The most precise locators of the FVIII binding site, however, are those mutations that are associated with a more severe type 2N phenotype; that is Glu787Lys (reported in only one patient thus far), Thr791Met and Arg816Trp[310–312,391] in TIL', all of which are largely buried (Figure 2.30 and Figure 2.42, upper panel). Surface charge mapping reveals that domain TIL' is an electric dipole, with an almost diagonal partitioning of positive and negative charge (Figure 2.42, lower panel). Two of the severe type 2N mutations, Thr791Met, and Arg816Trp (which occurs at a position of strong charge conservation - Figure 2.2), together with Thr789Pro, Met800Val, Arg816Gln and His817Gln, circumscribe a pocket of concentrated positive charge density on the front face of TIL' (Figure 2.42, lower panel), that is congruent with the burial of TIL' residue

clusters; that is, the pocket is bound laterally by strand β2 on the left and by an ascending segment of the β1-to-β2 loop on the right, superiorly, the pocket is enclosed by strand β4 and the upper portion of the β1-to-β2 loop (Figure 2.30 and Figure 2.42).

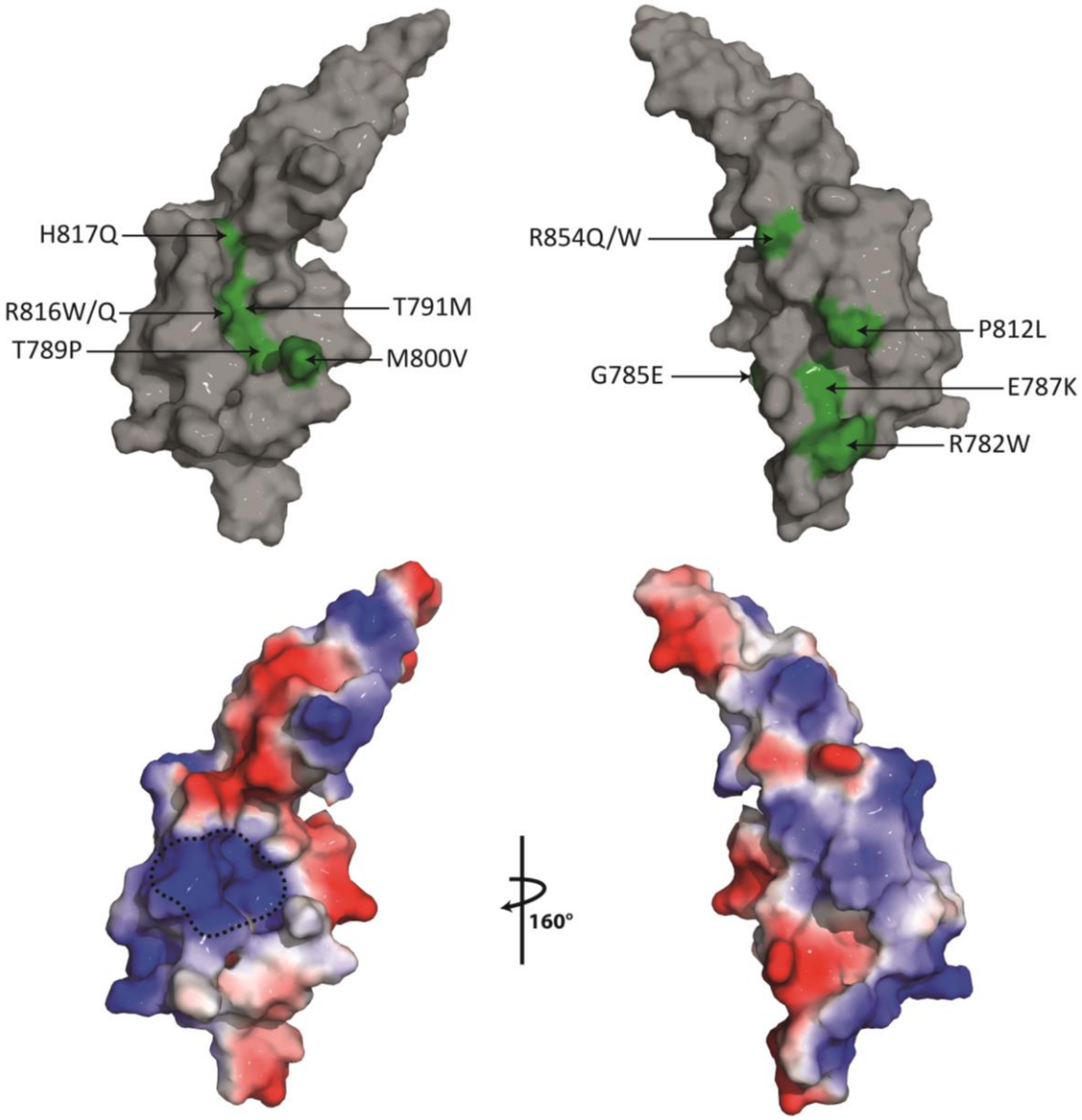


Figure 2.42: Structural interpretation of VWD type 2N mutants in TIL'E'. The structures on the right are rotated 160° about the y-axis, relative to structures on the left. The structures on the left show the front face of TIL'. Top panel; a surface representation of TIL'E' with positions of non-cysteine type 2N VWD mutations coloured green. Lower panel; electrostatic surface potential of TIL'E' is shown (blue, negative charge; red, positive charge).

The FVIII light chain N-terminal acidic a3 domain mediates high affinity binding to VWF, and it is thus, highly likely that the positively charged pocket on the front face of TIL' is the binding site for the negatively charged a3 domain. Consequently, mutants

Thr789Pro, Thr791Met, Met800Val, Arg816Trp/Gln and His817Gln mediate their effects through perturbation of the correct configuration of this dynamic positively charged pocket. In addition, whilst clinically milder substitutions Thr789Pro, Met800Val, Arg816Gln and His817Gln result in a decreased residue volume, the severity of Thr791Met and Arg816Trp mutants are accounted for by the introduction large side chains that would conceivably hinder accommodation of the FVIII light chain a3 domain (Table 2.5). Moreover, loss of positive charge mediated my mutants Arg816Gln/Trp and His817Gln only compounds the deleterious effect on FVIII binding.

Residues Arg782, Gly785 and Glu787 are in close proximity to this putative binding pocket (Figure 2.41 and Figure 2.42, upper panel), which suggests that the severe type 2N mutation Glu787Lys and mutations Arg782Trp and Gly785Glu facilitate the FVIII binding defect through charge disruptions and perturbation of the β1-to-β2 loop, the ascending and upper portions of which form the left superolateral border of the positively charged TIL' pocket (Figure 2.41). An effect of the consecutive TIL' Pro811 and Pro812 residues in the loop to strand-β3 confer conformational stiffness and introduce a turn that anchors strand β3. This effect serves to maintain the correct conformation of the positively charged TIL' pocket, wherein, substitution Pro812Leu may perturb the binding pocket through a degree of unwanted flexibility (Figure 2.41).

Although mutant Arg854Gln is clinically mild, the Arg854Trp variant causes a type 2N diathesis of moderately severe bleeding that, similarly to the type 2N Cys mutants, is associated with defective secretion and a significant abnormality of multimerisation, with lack of intermediate and high molecular weight VWF multimers[392]. The discrepancy between substitution to Gln or Trp at position 854 can be reasoned to the proximity of Arg854 in the E' β4-to-β5 loop to the TIL' disulphide bridge Cys792-Cys827 (linkage 4-10). Replacement of a positively charged linear arginine with the bulky hydrophobic ring of tryptophan at an exposed position (Δ +42.1 Å^3, Table 2.5) is predicted to cause significant structural perturbation of the C-terminal TIL' region, residues Leu824 to Cys827, encompassing the 3_{10} helix and bridge Cys792-Cys827, which anchors the upper portion of the TIL' β1-to-β2 loop to the TIL' C-terminus. This region borders the superior aspect of the positively charged putative FVIII a3 domain binding pocket in TIL', to which the effects induced by mutant Arg854Trp are likely to

be propagated. In addition, Trp at this position may threaten the linkage of Cys792 to Cys827 that would in turn, destabilise the ascending segment of the TIL' β1-to-β2 loop that borders the positively charged pocket. These effects would act in concert to account for the reduced FVIII binding and multimerisation defect observed in mutant Arg854Trp. In contrast, evolution of a Δ –40.7 Å^3 cavity by substitution Arg854Gln (Table 2.5), would not be expected to cause such destructive structural alterations to TIL', and hence, results in a milder phenotype, as E' most probably contributes to FVIII binding through subsidiary contacts, whilst TIL' is the principal FVIII binding site, by means of a3 binding.

Table 2.5: Type 2N VWD mutations in TIL'E'. Residue volumes are those of Pontius *et al*[393].

Domain	Mutation	Δ Volume (Å^3)	Position	Predicted effect
TIL'	R782W	+41.1	β1-β2 loop	Loss +ve, polar-to-hydrophobic, structural perturbation
	G785E	+88.1	β1-β2 loop	Introduce –ve charge, surface charge disruption, ↑volume
	E787K	+8.9	β1-β2 loop	Introduce +ve in region of –ve charge, highly deleterious
	C788R	+94.7	β1-β2 loop	Apparition of –SH, introduce +ve, ↑volume
	C788Y	+108.4	β1-β2 loop	Apparition of –SH, ↑volume
	T789P	–3.6	β1-β2 loop	Perturbation of putative a3 binding pocket
	T791M	+40.9	β1-β2 loop	Perturbation of putative a3 binding pocket
	Y795C	–96.4	β1-β2 loop	Exposed –SH, ↑volume
	M800V	–28.5	β1-β2 loop	Border of putative a3 binding pocket
	C804F	+85.4	β1-β2 loop	Apparition of –SH
	P812L	+41.0	β2-β3 loop	Affects surface contour of putative a3 binding pocket
	R816W	+42.1	β3	Loss of conserved +ve in putative a3 binding pocket
	R816Q	–40.7	β3	Loss of conserved +ve in putative a3 binding pocket
	H817Q	–7.8	β3	Loss of conserved +ve in putative a3 binding pocket
E'	R854W	+42.1	β4-β5 loop	Loss +ve, large aromatic ring
	R854Q	–40.7	β4-β5 loop	Loss +ve, creates cavity
	C858F	+97.4	β5	Apparition of –SH, steric bulk
	C858S	-0.4	β5	Apparition of –SH

2.15. Conclusion of results

In conclusion, most of the non-cysteine type 2N VWD mutations are clustered within dynamic regions of domain TIL' and their effects converge on perturbation of the surface contour of the positively charged TIL' pocket on the front face of TIL'. This localises the principal FVIII binding site, mediated by acidic FVIII domain a3, to this region of the VWF protein. A predominantly electrostatic interaction between VWF and the FVIII light chain is in agreement with the known dissociative effects of high

ionic strength and low pH on the VWF·FVIII complex. Hence, based on the foregoing analyses, it is proposed that the E' domain and the β-sheets of TIL' provide a rigid scaffold for the positively charged pocket of TIL', which complements the charge distribution of the FVIII light chain a3 domain; thus electrostatic steering is the dominant force that orientates, with high specificity, the VWF and FVIII interfaces (Figure 2.43). The final stages of binding may then proceed via conformational selection of the excited TIL' state, in which the TIL' binding pocket concavity complements the binding surface of the a3 domain on FVIII.

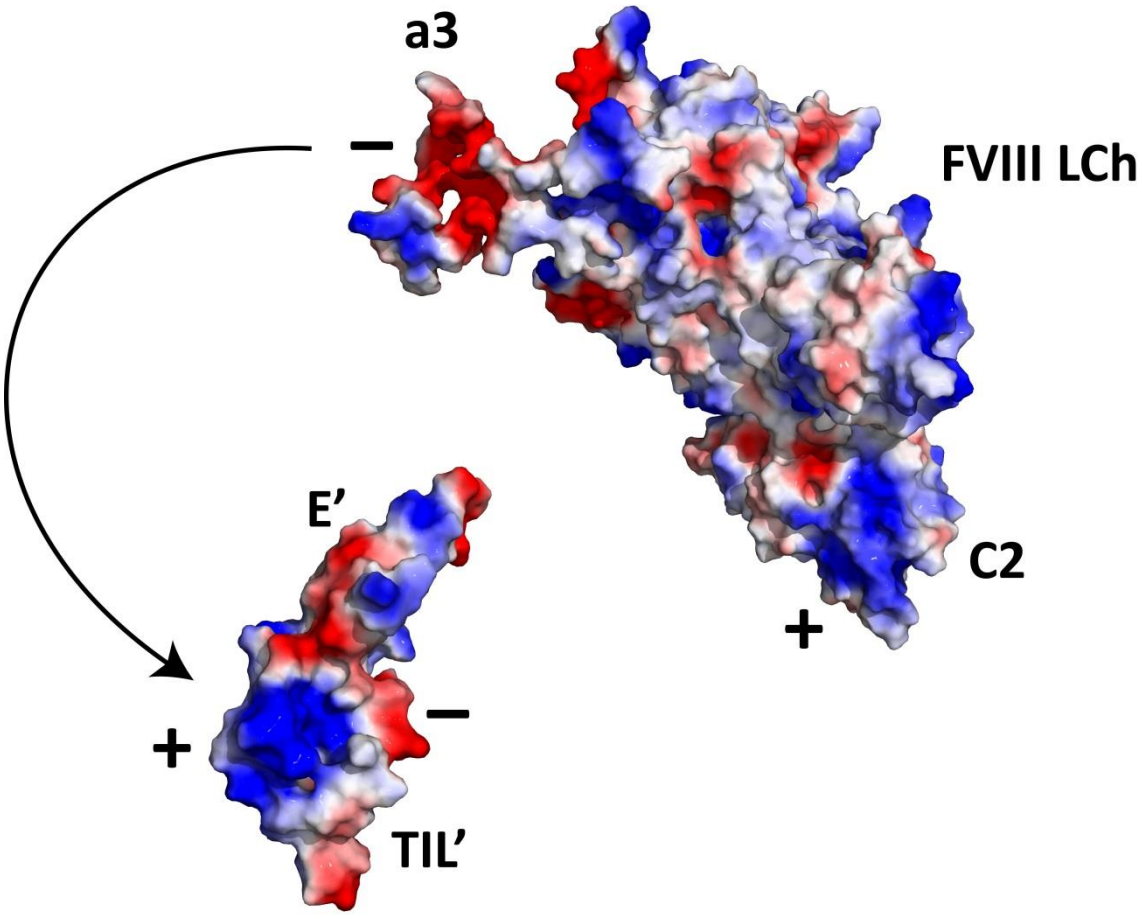


Figure 2.43: Mechanism of electrostatic VWF·FVIII binding. Shown is the proposed mechanism of high affinity FVIII light chain (LCh) a3 domain binding to VWF TIL' domain, by electrostatic steering forces.

3. Concluding remarks

The fascinating mechanisms by which VWF maintains haemodynamic integrity have enthused scientific interest since 1926 when a consanguineous family from the Aland archipelago of Finland, affected by a curious bleeding disorder, came to the attention of Dr Erik von Willebrand[394,395]. Much of the function of the eponymous VWF protein has since been elucidated and knowledge of the epidemiology, genetics and molecular basis of VWD is rapidly expanding. Indeed, extensive clinical observations and laboratory studies have revealed that VWF is unique and unusual, both in terms of its multimeric structure and the mechanisms governing its participation in haemostatic processes. Furthermore, VWF is implicated in pathological processes beyond haemostasis, including angiogenesis, inflammation and tumour cell survival[204].

Most intriguingly however, is the capacity of VWF to undergo both domain-level and gross conformational transitions in regions of elevated rheologic shear forces[138,396]. The interaction sites of circulating (globular) VWF are buried and therefore, inaccessible to their ligands; allowing VWF to conduct a circulatory reconnaissance of the intact vasculature. The functional quiescence of VWF is perturbed by vascular damage, exposing subendothelial collagens, to which globular VWF is recruited. Subsequently, VWF undergoes structural transitions mediated through the local shear forces exerted on the tethered VWF molecule by flowing blood[138], thereby, unravelling VWF into an elongated, haemostatically potent, biopolymer. Cryptic binding sites are thus revealed, and facilitate the capture and localisation of platelets to the denuded vessel wall. These first steps in the genesis of the primary platelet clot, as well as the transition of VWF from a paracrystalline highly ordered configuration when stored in WPBs to a disordered configuration in circulation, demonstrate that the conformation and dynamic malleability of VWF is fundamental to its haemostatic function[71].

Notwithstanding the significant advances in the characterisation of VWF function, little is known of the molecular structure that is critical to maintaining the delicate balance between bleeding and thrombosis. In particular, atomic-level insight into the VWF region that mediates FVIII stabilisation in plasma, is lacking. With the exception of the triplicated globular A domains[62,63,190,194–197], structural studies on VWF have been

limited to low-resolution measurements that contribute to understanding the overall shape and flexibility of the molecule. Early metal-shadowing electron microscopy experiments[137,139] revealed the head-to-head and tail-to-tail arrangement of VWF monomers within multimeric assemblies. Subsequent higher resolution negative stain electron microscopy studies have refined the overall size and arrangement of domains within VWF[140] (Figure 3.1).

A complete description of VWF function is dependent upon determination of atomic-resolution three-dimensional domain structures. Crystallisation of the VWF A2 domain[190] revealed the mechanism by which A2 unfolds to expose the ADAMTS13 cleavage site in the regulation of VWF multimer size and thrombogenicity[188,189]. The folded VWF A2 domain adopts an atypical Rossman fold, whereby amphipathic α-helices surround a central β-sheet, in which the ADAMTS13 Tyr1605-Met1606 scissile bond is deeply buried[190]. In place of an α4-helix, present in domains A1 and A3, the A2 domain contains a flexible α4-less loop that weakens the fold and reduces the force required for unfolding[172,190]. This structural instability is further compounded by the absence of a long-range N- to C-terminal disulphide bond, present in the neighbouring homologous A domains, that would otherwise protect the domain against unfolding by the shear forces in circulation[190,397]. The A2 crystal structure has also revealed a very rare vicinal disulphide bond between adjacent Cys1669 and Cys1670 at the C-terminus that acts as a molecular plug to impart rigidity and stabilise the hydrophobic core[190,398]. In regions of elevated shear flow however, the vicinal disulphide bond is pulled out, destabilising the A2 domain core and initiating unfolding to reveal the Tyr1605-Met1606 scissile bond and expose the VWF Glu1660-Arg1668 cryptic exosite[71]. Consequently, the VWF A2 domain becomes susceptible to proteolysis by ADAMTS13[71,172,190]. Thus, the evolutionary structural adaptations of the VWF A2 domain, revealed only when the high resolution structure was determined, predict that reversible unfolding and refolding regulate VWF cleavage by dynamic accessibility of a buried scissile bond[190,397].

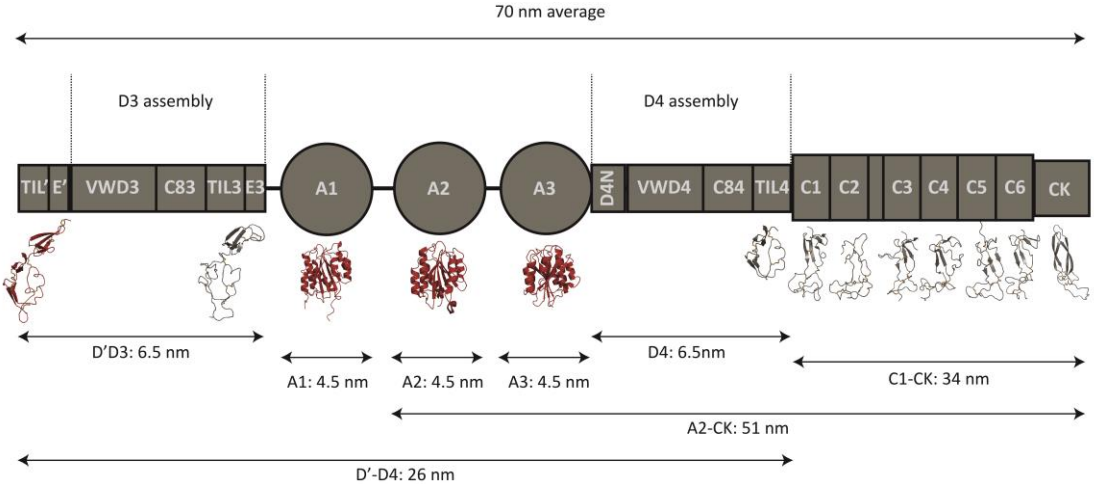


Figure 3.1: Overall dimensions of monomeric VWF. Shown is the domain architecture of mature monomeric VWF and the approximate lengths obtained by electron microscopy studies[137,139,140]. High resolution structures are shown below the domain schematic and are coloured red; these are the NMR structure of TIL'E' (D') and crystal structures of domains A1, A2 and A3. Structures shown in grey are homology models, of which modelling of TIL3E3 and TIL4 is based on the NMR structure of TIL'E'. The VWC and CK domains are modelled on pdb structures 1U5M[118] (VWC domain of human collagen IIA) and 1XWD (CK domain of human follicle stimulating hormone)[407], respectively. High resolution NMR and crystal structures reveal five functionally significant VWF domains (TIL', E', A1, A2 and A3) and homology predicts the majority of the remaining mature protein, with the exception of the VWD and C8 domains.

It is thus evident that the interplay between structure and dynamics regulates both the function of the VWF protein and its ability to evolve and adapt to diverse sets of conditions necessary for maintaining haemostasis[385]. At the outset of the work described in this thesis, the molecular mechanisms governing formation of the biologically important VWF·FVIII complex were unresolved. To reveal the structural adaptations of the FVIII binding region of VWF, the high resolution NMR structure of the D' region was solved and its conformational dynamism characterised.

Expression of TIL'E' (D') in *E. coli* and its subsequent purification to homogeneity, revealed that the monomeric TIL'E' fragment, but not a further truncation to domain TIL', represents the minimal folded VWF unit required to bind FVIII. This finding, together with the high resolution solution structure of TIL'E', may have important implications for the development of improved haemophilia A therapeutics that rely on prolonging the half-life of FVIII *in vivo*.

The VWF D' region is composed of two independently folding domains: TIL' and E', whose folds are dominated by structural disulphide bridges. The topologically complex

TIL'E' disulphide bond connectivity was unambiguously determined. This newly assigns the Cys788-Cys799 (3-5), Cys792-Cys827 (4-10) disulphide bonds in domain TIL' and all three disulphide bonds in domain E', Cys829-Cys581 (1-4), Cys846-Cys863 (2-6) and Cys849-Cys858 (3-5), and confirms the three TIL' Cys767-Cys808 (1-7), Cys776-Cys804 (2-6) and Cys810-Cys821 (8-9) disulphide bonds determined biochemically[42]. Both TIL' and E' are of the small disulphide-rich, all-β sheet class of proteins; the TIL' domain is formed of two short β-sheets and a long imposing loop that connects strands β1 to β2, whereas, the E' domain is largely formed of a triple-stranded all anti-parallel β-sheet.

Domain TIL' is strikingly dynamic and harbours a cluster of pathologic mutations that lead to decreased FVIII binding affinity (type 2N VWD). The major FVIII binding site on VWF is localised to a flexible, positively charged pocket on the front face of the TIL' domain; a region that is perturbed by the majority of 2N VWD mutations. Surface-charge mapping of the TIL'E' NMR structure reveals a potential mechanism by which the strongly and oppositely charged VWF TIL' pocket and FVIII a3 domain form a high affinity encounter-complex. Together, these findings describe the structural specialisation of VWF TIL'E' in mediating FVIII binding and stabilisation in plasma, providing insight into the molecular basis of type 2N VWD.

The FVIII binding site is contained within domains TIL'E'VWD3 (residues 767 to 1031). High affinity binding of the FVIII light chain acidic a3 domain to VWF is very likely to be mediated by the positively charged pocket of TIL'. However, the contributions to binding by the VWF VWD3 domain and the FVIII light chain domains C1 and C2, are yet to be determined. The next logical step in this study is the biochemical characterisation of the binding affinity of the FVIII light chain to VWF TIL'E' and VWD3 domains. This would reveal differential binding affinity between TIL'E' and VWD3 to the FVIII light chain and test the hypothesis that high affinity binding is mediated by domains TIL'E'. Ideally, FVIII binding studies will also include the full TIL'E'VWD3 region, to account for cooperativity between the domains in formation of the binding interface. However, recombinant expression of this 24 cysteine fragment in *E. coli*, will likely pose a significant challenge, whereas the eight cysteine VWD3 domain has already been successfully expressed.

Following characterisation of the VWF TIL'E' and VWD3 binding kinetics, it would be of interest to determine the NMR structure of the VWD3 domain. This would reveal the entire FVIII binding region on VWF. Knowledge of the TIL'E' and VWD3 structures may then be combined with NMR binding studies, to identify the precise TIL'E'VWD3 binding face by mapping of contact residues in TIL'E' and VWD3. This would confirm if, as suspected, TIL'E' is the principal FVIII binding region on VWF, with subsidiary contacts to domain VWD3. Chemical shifts of residues at the interface of the TIL'E'·FVIII or VWD3·FVIII complex are likely to differ between the bound and free forms. The perturbation of the chemical shift upon complex formation can be used to map the interaction interface, and this data is subsequently used to restrain docking of the FVIII light chain onto the TIL'E' and VWD3 NMR structures. In essence, the ^{15}N-HSQC spectrum is monitored when unlabelled FVIII light chain is titrated into the isotopically labelled VWF domain. These studies may be complemented by NMR cross-saturation experiments and by TIL'E'·FVIII co-crystallisation trials, that would facilitate the rigorous identification of contact residues in the complex. The dependence of the TIL'E' and VWD interactions with FVIII on ionic strength and the effects of important type 2N missense mutations on complex formation may then also be studied in this manner. Type 2N VWD mutants in domains TIL'E' have been expressed for biochemical and NMR studies to characterise changes in both the FVIII binding affinity and in the TIL'E' structure and dynamics.

The most immediate succeeding steps are those of finer characterisation of the VWF·FVIII interaction, and the structural and dynamic perturbations upon inheritance of a type 2N VWD mutation, to build a detailed mechanistic model of VWF·FVIII complex formation. However, a critical element of the work presented in this thesis has been the successful expression and purification of TIL'E' in *E. coli*. This methodology forms a basis from which the recombinant production of the remaining cysteine-rich N- and C-terminal flanking regions may be optimised, leading to their structural characterisation. Of particular interest is the characterisation of the VWD and C8 domains, for which there are no known structural homologues. This would allow a high resolution structure of mature VWF to subsequently be assembled. In

view of the importance of VWF in common vascular diseases, its molecular structure is not only of basic scientific interest but is also of direct clinical importance.

4. Materials and Methods

The methodologies for the studies presented in this thesis are outlined below.

4.1. Multiple sequence alignments of TIL' and E' domain families

The classical VWF D domains are annotated as containing VWD domains, C8 domains and trypsin inhibitor-like (TIL) domains in the protein family, Pfam[399], database (VWD: PF00094; C8: PF08742; TIL: PF01826) and VWD and TIL domains in the universal protein source (UniProt; human VWF: P04275.4). The E domains, assigned in this study and by Zhou YF *et al*[50], follow the VWF TIL domains. The Pfam database[399] annotates a fibronectin I-like clan (CL0451) based on sequence homology that includes VWC domains. Their structural homology is corroborated by the structural classification of proteins (SCOP) database[400] in which VWC subdomain 1 and fibronectin are classified to the fibronectin I-like superfamily (57603). These databases are yet to include the VWF E domains in this superfamily.

VWF (NP_000543.2) and homologous domain family sequence alignments were initially built using the programme MUSCLE[401]. These alignments were subsequently viewed and refined by eye, to account for cysteine conservation and hydrophobic residues, using the programme Jalview[402]; these multiple sequence alignments were used to define the VWF TIL' and E' domain boundaries.

4.2. Construction of homology models

The final structural templates, used to model domains TIL' and E', were principally based on two parameters; the highest sequence identity (similarity) and the least number of insertions and deletions in a pairwise alignment. Sequence identity (and similarity) was assessed by global pairwise alignment using the programme Needle (EMBOSS)[114]. VWF TIL' was modelled against the NMR structure of *Apis mellifera* chymotrypsin/cathepsin G inhibitor-1 (AMCI-1; pdb: 1CCV), sharing 32.3% sequence indentity (37.1% similarity). Modelling of TIL' based on this template required two

insertions, one of eight resides and another of one residue, and two single residue deletions. Domain E' was modelled against the NMR structure of VWC subdomain 1 of human collagen IIA 30.6% (similarity 38.9%), requiring one single residue deletion. To induce Cys_2-Cys_6 bonding in domain E', Cys_6 was aligned to Cys_8 of template 1U5M. Dictionary of Secondary Structure in Proteins (DSSP)[403] assignments of the template structures were included into the pairwise alignments to correctly assign sequence gaps to loop regions, and to conserve the positions of the cysteine residues. Models were subsequently built and energy minimised with the programme Modeller 9v8 using the automodel class[404].

4.3. Cloning of TIL'E' into pET32b(+) vector - construct TRX-His_6-TIL'E'

Plasmid pcDNA 3.1 FL-VWF containing the full length VWF cDNA was a kind gift from Dr Thomas AJ McKinnon. The VWF coding sequence corresponding to domains TIL'E', residues 766-864 (one end residue), contains *XhoI* and *NcoI* restriction sites. These sites were removed by conservative site directed mutagenesis to facilitate cloning of this region into vector pET32b(+).

The VWF TIL'E' domains (residues 766-864) span bases 2296 to 2592 (297 bp) and contain an *NcoI* site at cDNA bp 2309-2314 (CCATGG) and an *XhoI* site at bp 2356-2361 (CTCGAG) in domain TIL'. The following complementary mutagenic oligonucleotides (orientated 5' to 3') were used to remove these restriction sites in successive mutagenesis reactions by the introduction of mutations c.2310C>G, to remove the ***NcoI*** site and c.2358C>G, to remove the ***XhoI*** site:

NcoI Forward: ATC CTG TCG GCC CCC **G**AT GGT CAA GCT GGT G (31 bp)
Reverse: CAC CAG CTT GAC CAT **C**GG GGG CCG ACA GGA T (31 bp)

XhoI Forward: CGG GCT GAA GGG CT**G** GAG TGT ACC AAA AC (29 bp)
Reverse: GTT TTG GTA CAC TC**C** AGC CCT TCA GCC CG (29 bp)

Mutagenesis of VWF cDNA was performed by the polymerase chain reaction (PCR), in a Corbin Research Thermal Cycler, using the (***NcoI*** and ***XhoI***) oligonucleotides outlined

above, and KOD Xtreme™ Hot Start DNA Polymerase (Novagen). The reaction set up is outlined below:

Table 4.1: KOD Xtreme polymerase reaction set up

Component	Stock	Volume	Final concentration
Xtreme buffer	2 x	25 μL	1 x
dNTPs	2mM each	10 μL	0.4 mM each
ddH_2O	-	10 μL	-
5' primer	10 μM	1.5 μL	0.3 μM
3' primer	10 μM	1.5 μL	0.3 μM
FL-VWF DNA template	378 ng/μL	1 μL	378 ng
KOD Xtreme DNA polymerase	1 U/μL	1 μL	0.02 U/μL
Total reaction volume		**50 μL**	

The following PCR temperature cycling protocol was used for mutagenesis:

Table 4.2: Mutagenic PCR temperature cycling

Step	Temperature (°C)	Time	Number of cycles
Polymerase activation	94	30 s	1
Denaturation	98	10 s	
Annealing	55	30 s	16
Extension*	68	13 min 46 s	
Completion	68	10 min	1
Hold	4	-	-

* Extension calculated at 1 minute per Kilo bp

Successful PCR reactions were verified by agarose gel electrophoresis. An 0.7% w/v agarose gel was made by dissolving 0.35 g agarose in 50 mL 1 x TAE buffer and heating. A stock of 50 x TAE buffer was used (242 g Tris, 57.1 mL glacial acetic acid, 100 mL 0.5 M EDTA at pH 8.0). The agarose solution was left to cool and 2.5 μL of ethidium bromide (10 mg/mL) was added to a final concentration of 0.5 μg/mL. The solution was poured into a gel frame containing a comb for well formation, and allowed to set at room temperature for 40 minutes. 5 μL of PCR reaction product was mixed with 1 μL of 6 x DNA loading buffer (Novagen) and applied into the well. The TrackIt™ 1 Kb Plus DNA ladder (Invitrogen) was used for reference, and the samples electrophoresed in 1

x TAE buffer at a constant voltage of 100V until the dye front had reached the end of the gel. The DNA bands were visualised under ultraviolet light by ethidium bromide fluorescence.

Selection of mutant VWF cDNA was achieved by treatment with *DpnI*. The *DpnI* endonuclease (target sequence: 5'-Gm^{6}ATC-3') is specific for methylated and hemimethylated DNA and is used to digest parental DNA template and to select for mutation containing synthesised DNA. Plasmid DNA isolated from almost all commonly used strains of *E. coli* have been fully methylated *in vivo* by the endogenous Dam methylase and are therefore sensitive to cleavage by *DpnI*. In contrast, DNA synthesised *in vitro* using the conventional four deoxynucleotides is unmethylated and therefore completely resistant to cleavage. Consequently, *DpnI* can be used after the site directed mutagenesis to degrade residual methylated wild-type templates and to enrich for unmethylated DNAs synthesised *in vitro*. The *DpnI* reaction was set up as follows using reagents from New England Biolabs (NEB):

Table 4.3: *DpnI* digestion reaction

Component	Volume (μL)
PCR product	40
BSA (100 x)	0.5
NEB Buffer 4 (10 x)	5
ddH_2O	3.5
DpnI (20 U/μL)	1

The *DpnI* reaction (Table 4.3) was incubated for two hours at 37 °C, and the endonuclease subsequently heat inactivated at 65°C for 20 minutes. The reaction was then placed on ice awaiting transformation into NovaBlue^{TM} (Novagen) competent cells for plasmid purification and sequencing, to confirm that both the *XhoI* and *NcoI* sites within the VWF TIL' have been mutated.

To 25 μL of *E. coli* NovaBlue^{TM} (Novagen) cells, 1 μL or 1.5 μL of *DpnI* digested mutated pcDNA 3.1 FL-VWF plasmid was added. The cells were incubated on ice for 30 minutes, followed by a heat shock of 45 seconds at 42°C and placed back on ice for 1 minute. 800 μL of sterile LB (Luria Broth) media was added to the cells and incubated for one

hour in a 37 °C orbital shaker at 200 rpm, to aid cell recovery and allow expression of the ampicillin resistance gene. 250 μL of cells were then plated onto LB agar plates containing 50 μg/mL ampicillin and incubated at 37 °C for 12-16 hours, until the colonies were 2-3 mm in size.

Single colonies were inoculated into 5 mL of sterile LB media containing 50 μg/mL ampicillin and grown overnight. DNA was subsequently purified using the E.Z.N.A plasmid miniprep kit (Omega Bio-Tek). The mutations generated to remove TIL' domain restriction sites were verified by sequencing (Source Bioscience) to ensure the absence of PCR-introduced errors and the presence of the desired mutations (c.2358C>G and c.2310C>G).

VWF cDNA bp 2296-2592 (297 bp) encompassing domains TIL'E' (residues 766 through 864) was subsequently amplified from the correctly mutated pcDNA 3.1 FL-VWF plasmid. The forward primer was designed to amplify VWF from base number 2296 and introduced 10 random bases followed by an *NcoI* restriction site before the start of the VWF sequence. TTT TTT CAT G random bases were inserted to reduce the GC content and melting temperature of the oligonucleotide; the last four bases of which, CAT G, also serve to increase *NcoI* specificity. Following the *NcoI* site, an additional two bases, GC, were added to keep the sequence in frame. The reverse primer amplified from VWF base number 2592 and, immediately following the VWF sequence, introduced a stop codon (TTA), a *XhoI* site and six random bases (CCG CCG), to increase *XhoI* specificity and the melting temperature of the primer. The oligonucleotides used, orientated 5' to 3' are outlined below, in which the restriction sites are shown in bold:

Forward: TTT TTT CAT G**CC ATG G**GC TCC TGT CGG CCC CCG ATG Tm: 75.16 °C

Reverse: CCG CCG **CTC GAG** TTA ATC ACA CAC ATG GTC TGT GCA Tm: 74.00 °C

The amplification reaction was set up as outlined in Table 4.1, and the PCR temperature cycling protocol used for DNA amplification is shown in Table 4.4. A 0.7% w/v agarose gel confirmed amplification of PCR products. The resulting VWF product was double digested with restriction enzymes *NcoI* and *XhoI*, and their respective buffers (NEB). The 5.9 Kb pET32b(+) cloning vector contains a T7 promoter, N-terminal

thioredoxin and His_6 tags, and a gene conferring ampicillin resistance. The vector has been modified to replace an enterokinase site with a Tobacco Etch Virus (TEV) cleavage site (ENLYFQ$^{\downarrow}$G). To facilitate cloning of the VWF amplification product into vector pET32b(+) using restriction sites *NcoI* and *XhoI*, the vector was also digested with *NcoI* and *XhoI* restriction enzymes (NEB). The restriction enzyme digestion reaction components are outlined in Table 4.5. Digestion was allowed to proceed at 37°C for three hours to produce the VWF TIL'E' insert and the linearised pET32b(+) vector.

Table 4.4: Amplification PCR temperature cycling

Step	Temperature (°C)	Time	Number of cycles
Polymerase activation	94	2 min	1
Denaturation	98	10 s	
Annealing†	74	30 s	30
Extension*	68	20 s	
Completion	68	1 min	1
Hold	4	-	-

†The annealing temperature is that of the lowest melting temperature (Tm) primer. * Total amplified DNA product is 330 bases, extension time calculated at 1Kb/minute giving 19.8 seconds.

Table 4.5: *XhoI* and *NcoI* double digestion reaction

Component	Volume (μL)
VWF amplified DNA or Pet32b(+) vector	20
XhoI	3
NcoI	3
BSA (100 x)	1
NEB buffer 4	10
ddH_2O	63
Total reaction	100

To purify the VWF TIL'E' insert and pET32b(+) vector DNA prior to ligation, the samples were subjected to 0.7% w/v agarose gel electrophoresis. The DNA fragments were visualised by ethidium bromide fluorescence upon exposure to ultra violet light, and excised from the gel. The insert and vector DNA were purified from the gel slices using an E.Z.N.A gel extraction kit (Omega Bio-Tek) and eluted in 30 μL elution buffer (10 mM Tris·HCL, pH 8.5) or water.

The VWF TIL'E' (aa 766-864) insert containing base pairs 2296 to 2592 and vector pET32b(+) were ligated using T4 DNA ligase and 10 x T4 DNA ligase buffer (NEB). Molar ratios of 1:3, 1:5 and 1:10 vector-to-VWF insert were used for cohesive end ligations. The amount of insert required for each molar ratio was determined using the equation below:

$$\frac{ng\ vector\ x\ size\ insert\ (kb)}{size\ vector\ (kb)}\ x\ molar\ ratio\ = ng\ insert$$

Ligation reactions were incubated at 16°C for 18 hours. T4 DNA ligase activity was subsequently quenched by heat inactivation for 10 minutes at 65°C. 10 μL of each ligation reaction was then transformed into 50 μL of NovaBlue^{TM} (Novagen) cells, plated on LB agar containing 50 μg/mL ampicillin, and incubated overnight at 37°C.

Single bacterial colonies from each plate were screened by PCR to confirm the presence of the VWF TIL'E' (bp 2296-2592; aa 766-864) insert using the VWF amplification (cloning) primers outlined above. Taq polymerase (NEB) was used for all PCR colony screening reactions. Sterile pipette tips were used to collect single colonies, which were mixed into a PCR tube containing the reaction components set in Table 4.6. Simultaneously, that colony was also inoculated into 5 mL of LB media containing 50 μg/mL ampicillin and grown overnight at 37°C in an orbital incubator at 200 rpm, for plasmid purification and sequencing upon indication of a positive PCR screen.

Table 4.6: Colony screen reaction conditions

Component	Volume (μL)
Taq ThermoPol buffer (10 x)	2.5
Forward VWF amplification primer (10 μM)	1.25
Reverse VWF amplification primer (10 μM)	1.25
dNTPs (2mM each)	2.5
Sterile ddH_2O	17
Taq DNA polymerase	0.5

The temperature cycling protocol used for the colony screen is outlined below:

Table 4.7: Colony screen PCR temperature cycling

Step	Temperature (°C)	Time	Number of cycles
Polymerase activation	95	5 min	1
Denaturation	95	1 min	
Annealing†	69	1 min	30
Extension	72	30 s	
Completion	68	1 min	1
Hold	4	-	-

† Tm of lowest primer (74°C) minus 5°C.

Two control PCR reactions were also set up. Negative control reactions contained no bacterial colony. Positive control reactions contained 0.5 μL of restriction site mutated pcDNA 3.1 FL-VWF. PCR products of the screen were analysed by agarose gel electrophoresis for successful amplification of the VWF TIL'E' insert.

Plasmids containing the VWF insert were purified from the overnight cultures using the E.Z.D.N.A plasmid miniprep kit (Omega Bio-Tek) and digested with *XhoI* and *NcoI* restriction enzymes (NEB) to confirm clones containing the VWF insert of the expected size (~300 bp). Correct clones were sent for sequencing (Source Bioscience). Construct TRX-His_6-TIL'E' (VWF bp 2296-2592; residues 766-864) with the correct sequence was used for protein expression, purification and TIL'E NMR structure determination.

To produce construct TRX-His_6-TIL' (VWF residues 766-828) a stop codon (TAA) was inserted following VWF Pro828 (CCC) in construct TRX-His_6-TIL'E' (VWF bp 2296-2592;

residues 766-864), using the complementary pair of mutagenic oligonucleotides shown below (inserted bases are highlighted in capitals):

Forward:	gcc ctg gaa agg tgt ccc TAA tgc ttc cat cag ggc aag	Tm: 76°C
Reverse:	Ctt gcc ctg atg gaa gca TTA ggg ac acct ttc cag ggc	Tm: 76°C

4.4. Cloning of TIL'E' into pET29b(+) - construct His_6-TIL'E'

Many difficulties were encountered in the generation of construct His_6-TIL'E' (VWF residues 766-864), for reasons which are unknown. Initially VWF sequence bp 2323 to 2592, encompassing residues 775 through 864 (domains TIL'E' but excluding residue 767 Cys_1 of TIL') was amplified from pcDNA3.1 FL-VWF plasmid containing mutation c.C2358G to remove the *XhoI* restriction site in domain TIL', and cloned into vector pET29b(+) containing an N-terminal His_6 tag and kanamycin resistance, using restriction sites *BamHI* and *XhoI* (NEB). Repeated attempts to clone VWF bp 2296-2592 (residues 766-864), containing the correct TIL'E' domain boundaries into pET29b(+) vector, were unsuccessful. Cloning oligonucleotides for the insertion of VWF sequences into pET29b(+) are shown below:

VWF bp 2323-2592 (residues 775-864)

Forward:	TTT TTT GGA TCC GTG TGT CCC GCT GAC	Tm: 66.5°C
Reverse:	TGG TTG CTC GAG TTA ATC ACA CAC ATG GTC	Tm: 66.8°C

VWF bp 2296-2592 (residues 766-864)

Forward:	TTT TTT GGA TCC TCC TGT CGG CCC CCC ATG	Tm: 70.9°C
Reverse:	TGG TTG CTC GAG TTA ATC ACA CAC ATG GTC	Tm: 66.8°C

Therefore, the VWF sequence (bp 2323-2592) corresponding to residues 755-864, cloned into vector pET29b(+), was extended by a pair of mutagenic oligonucleotides designed to insert the missing nine VWF residues (766-774). This resulted in the

construct His_6-TIL'E' (VWF bp 2296-2592; residues 766-864), that encompasses all conserved TIL' cysteines. The mutagenic nine-residue insertion primers are shown below; inserted base pairs are capitalised.

Forward: ctg tac ttc cag gga tcc TCC TGT CGG CCC CCC ATG GTC AAG CTG gtg tgt ccc gct gac aac Tm: 84 °C

Reverse: gtt gtc agc ggg aca cac CAG CTT GAC CAT GGG GGG CCG ACA GGA gga tcc ctg gaa gta cag Tm: 84 °C

These primers were successfully used with KOD Xtreme™ Hot Start DNA Polymerase in a reaction setup as outlined in Table 4.1, using the temperature cycling protocol shown in Table 4.8. This facilitated the extension of the VWF sequence in pET29b(+) from bp 2323-2592 (residues 775-864) to bp 2296-2592 (766-864), forming construct His_6-TIL'E', which was eliminated from further study as expression produced disulphide isomers of incorrectly folded TIL'E' molecules.

Table 4.8: Nine-residue insertion mutagenisis PCR reaction

Step	Temperature (°C)	Time	Number of cycles
Polymerase activation	94	30 s	1
Denaturation	98	10 s	
Annealing	55	30 s	18
Extension*	68	5 min 41 s	
Completion	68	10 min	1
Hold	4	-	-

* calculated at 1 Kb/minute

4.5. Expression of TRX-His_6-TIL'E' in *E. coli* SHuffle cells

Construct TRX-His_6-TIL'E' (residues 766-864; bp 2296-2592) cloned into the *E. coli* pET32b(+) vector containing an N-terminal thioredoxin (TRX) and His_6 tag was used for all further studies. The TRX domain aids the formation of eight disulphide bonds and the His_6 tag allows purification by nickel affinity chromatography (Ni-NTA) - was used for all further studies. The TIL'E' (VWF residues 766-864) fusion protein was expressed

in SHuffle cells[337] and grown in either LB media or M9 minimal media. An LB agar plate containing 50 μg/mL ampicillin was inoculated with SHuffle cells[337], that had been transformed with plasmid pET32b(+) TRX-His_6-TIL'E' and grown overnight at 37 °C.

Isotope labelled TIL'E' for NMR studies, was grown in M9 minimal media supplemented with 2.5 g/L ^{13}C-D-glucose and 1 g/L $^{15}NH_4Cl$, as the sole source of carbon and nitrogen, respectively. For single isotope labelling, $^{15}NH_4Cl$ was substituted for NH_4Cl and ^{13}C-D-glucose was substituted for D-glucose (10 g/L). Components of 1L of M9 minimal media are outlined below:

M9 base pH 7.4	1 L
$^{15}NH_4Cl$	1 g
^{13}C-D-glucose	2.5 g
1 M $MgSO_4$	1 mL
0.1 M $CaCl_2$	1 mL
Micronutrient solution 1 (1000 x)	1 mL
Micronutrient solution 2 (10 000 x)	0.1 mL
Thiamin/biotin solution (1000 x)	1 mL
Ampicillin (1000 x)	1 mL

M9 base:

Na_2HPO_4	6 g
KH_2PO_4	3 g
NaCl	0.5 g

Adjusted to pH 7.4 and autoclaved

Micronutrient solution 1 (1000 x)

$CaCl_2$	1.5 g
Na_2EDTA	3 g
$FeCl_3$	2.5 g

Micronutrient solution 2 (10 000 x)

$CuSO_4$	2.4 g
$MnSO_4$	1.8 g
$ZnSO_4$	0.27 g
$CoCl_2$	0.27 g

Thiamin/biotin solution (1000 x)

Thiamin	0.1 g
D-Biotin	0.1 g

Isotopically labelled cell cultures were initiated by inoculating 50 mL of M9 minimal media with a single transformed bacterial colony and grown overnight in an orbital shaker (220 rpm) at 30^{o}C to an OD_{600} of ~0.5.

The following morning, cells of the overnight culture were gently pelleted by centrifugation for 7 minutes at 4000 rpm. The cells were resuspended in 8 mL of fresh M9 media, which was used to inoculate 800 mL M9 media in a 2L flask (often 3.2 L of isotope labelled TIL'E' was grown). Cells were subsequently grown by shaking (200 rpm) at 30^{o}C to an OD_{600} of 0.4; after which time, the temperature was lowered to 16^{o}C and the cells were grown to an OD_{600} of 0.6. At this point, over-expression of TRX-His_6-TIL'E' was induced by the addition of 0.4 mM IPTG and cells were incubated for a further 16 hours at 16^{o}C. The cells were harvested by centrifugation (4000 rpm, 20 minutes, 4°C) and subsequently flash frozen in liquid nitrogen and stored at -80°C prior to lysis.

For 10% fractionally ^{13}C-labelled TRX-His_6-TIL'E', the M9 media was supplemented with 10% ^{13}C-D-glucose and 90% unlabelled glucose to drive the stereospecific labelling of leucine and valine methyl groups. Unlabelled TRX-His_6-TIL'E' for biochemical studies was grown in LB media containing 50 μg/mL ampicillin.

4.6. Purification of TIL'E'

TRX-His_6-TIL'E' cell pellets were thawed on ice and resuspended in 35 mL of wash buffer (20 mM sodium phosphate, 500 mM NaCl, 40 mM Imidazole, pH 7.4) by vortexing without pause to prevent the formation of foam. The cell suspension was then spiked with: 1 tablet of Complete protease inhibitor cocktail (Roche), 0.25% of the anionic detergent IGEPAL, lysozyme to a final concentration of 0.01 mg/mL and DNase (~5-10 μg/mL). The cells were subsequently broken by triple-pass through a French press and incubated for 30 minutes, under continuous rocking at 4°C, to complete lysis. Cellular debris and nuclei were removed by centrifugation at 20 000 rpm in a Sorvall SS34 rotor for 40 minutes.

The cleared SHuffle cell[337] cytoplasm is then filtered through an 0.22 μm syringe filter and loaded onto a His-Trap Ni-NTA column (5 mL; GE Healthcare) pre-equilibrated with wash buffer. The bound TRX-His_6-TIL'E' protein (28.6 kDa) was eluted using an AKTA FPLC system (Amersham) by the application of a shallow (20 column volume) linear gradient of increasing imidazole (0-500 mM) with the following elution buffer: 20 mM sodium phosphate, 500 mM NaCl, 40 mM, pH 7.4. Bound TRX-His_6-TIL'E' elutes at an imidazole concentration of 100-150 mM. Elution fractions were analysed by 4-20% Tris-glycine SDS-PAGE under non-reducing conditions. Fractions containing monomeric TRX-His_6-TIL'E' proteins are pooled (~20 mL) and dialysed against 20 mM sodium phosphate, 100 mM NaCl, pH 7.4, to reduce the imidazole concentration and reach optimal conditions for TEV proteolysis.

His-tagged TEV protease (~21 kDa) is stable when stored in 30% glycerol, 20 mM monobasic phosphate buffer, 250 mM NaCl and 10 mM β-mercaptoethanol. Crucially, the reducing agent β-mercaptoethanol is removed by dialysis against 20 mM sodium phosphate, 100 mM NaCl, pH 7.4, prior to proceeding with proteolysis of TRX-His_6-TIL'E'. For the TEV cleavage reaction, TEV protease is added to TRX-His_6-TIL'E' in the ratio of 1 OD_{280} TEV per 5-10 OD_{280} of TRX-His_6-TIL'E'; the reaction was incubated on a gently rocking platform at room temperature for three hours, and then at 4°C overnight to complete cleavage. Most of the TRX-His_6-TIL'E' protein, however, is cleaved within the first two hours.

Following cleavage, imidazole is added to the cleavage reaction to a final concentration of 40 mM, and subsequently applied again to the His-Trap Ni-NTA column (5 mL; GE Healthcare), pre-equilibrated with wash buffer (20 mM sodium phosphate, 500 mM NaCl, 40 mM Imidazole, pH 7.4). The N-terminal TRX-His_6 tag (17.3 kDa) binds to the nickel column, whereas TIL'E' remains unbound, and is thus collected in the binding and washing flow-through fractions. TEV-cleaved TIL'E' is formed of 103 residues, the amino-terminal four residues (GAMG) are artefacts of cloning. The TIL'E' pool is then concentrated and loaded onto a 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated with running buffer, 20 mM, as a final purification step. TIL'E' consistently elutes in a single symmetrical peak at 73 mL. Non-reducing SDS-PAGE (4-20% Tris-glycine) was used to analyse the size exclusion chromatography elution fractions.

Monomeric TIL'E' (11.3 kDa) in 20 mM sodium phosphate, 100 mM NaCl, pH 7.4 is concentrated to 450 μL with final sample concentrations obtained ranging from 0.5-1 mM. Protein concentration was determined by measurement of the UV absorbance at 280 nm, using an extinction coefficient of 0.836 for a 0.1% (mg/mL) solution. TIL'E' in the absence of an N-terminal thioredoxin tag and thioredoxin-tagged domain TIL' (residues 766-828) were expressed and purified in essentially the same way.

4.7. Ellman's assay

Free thiol groups in purified TIL'E' were quantified by molar absorptivity. That all cysteine residues were engaged in disulphide bonding was verified by an Ellman's assay (DTNB, 5,5'-dithio-*bis*-(2-nitrobenzoic acid)). DTNB reacts with a free sulfhydryl group to yield a mixed disulphide bond and TNB (2-nitro-5-thiobenzoic acid), which can be detected spectrophotometrically at 412 nm. The thiol content of purified monomeric TIL'E' was zero, as demonstrated by non-absorbance at 412 nm during the assay. Ellman's reagent was purchased from Thermo Scientific.

4.8. Nanoelectrospray ionisation mass spectrometry

Recombinant TIL'E' (10 μM) was buffer exchanged into 100 mM ammonium acetate (pH 7.0) for mass spectrometry analysis on a Synapt HDMS (Waters) mass spectrometer. A 2.5 μL aliquot of TIL'E' was delivered to the mass spectrometer by means of nanoelectrospray ionisation (nanoESI) using gold-coated capillaries. The experiments were kindly performed by Dr Konstantinos Thalassinos.

4.9. Blotting with recombinant FVIII

Binding of TRX-His_6-TIL'E' to full length FVIII was shown by dot blot analysis. Reconstituted recombinant FVIII, Helixate NextGen 0.25 μg/IU (CSL Behring), was buffer exchanged into 20 mM Sodium Phosphate, 100 mM NaCl, pH 7.4 to a concentration of 1 μg/μL. FVIII (125 ng, 250 ng, 500 ng and 1 μg) was blotted onto a nitrocellulose membrane and left to dry at room temperature for 1 hour. The

membrane was washed twice in TBS buffer (10 mM Tris.Cl, 100 mM NaCl, pH 7.5) at room temperature and subsequently blocked with 5% non-fat milk in TBS containing 0.05% tween 20 (TBS-T). This was followed by three wash steps, twice with TBS-T and a final wash with TBS. The membrane was then incubated for two hours at room temperature in blocking buffer (5% non-fat milk in TBS-T) containing 0.79 mg/mL TRX-His_6-TIL'E'. Following washing of the membrane, twice in TBS-T and once in TBS, it was subsequently incubated with horseradish peroxidase (HRP)-conjugated anti-(penta)His antibodies (Qiagen), diluted 1 in 1000 in blocking buffer, for one hour at room temperature. The membrane was washed again, and the binding of TRX-His_6-TIL'E' to the immobilised FVIII was revealed by ECL chemiluminescence (SuperSignal West Pico Chemiluminescent Substrate, Thermo Scientific). A control blot in which FVIII (62.5 ng, 125 ng, 250 ng, 500 ng and 1 μL), TRX-His_6-TIL'E' (30 ng) and BSA (250 ng) were spotted onto the membrane and detected with HRP-conjugated anti-His antibodies, was generated by the same procedure; importantly however, the control blot was not incubated with TRX-His_6-TIL'E' in blocking buffer.

4.10. Blotting with mAb 418

mAb 418 is a potent FVIII binding-blocking antibody that recognises the unreduced N-terminus of von Willebrand factor (epitope residues 765-816)[237,313]. An aliquot of mouse mAb 418 (0.65 mg/mL) was a kind gift from Dr Peter J Lenting. TIL'E' (0.45 mg/mL) was reduced with 10 mM dithiothreitol (DTT) and subsequently alkylated in the dark with a molar excess (30 mM) of iodoacetamide. 1 μL of a 1 M DTT solution was added to 100 μL of TIL'E' (10 mM final DTT concentration) and incubated on the bench (at room temperature) for 30 minutes. To the reduced TIL'E', 1.1 μL of 300 mM iodoacetamide was added to a final concentration of 30 mM; to alkylate and block the newly formed TIL'E' thiol groups, and thereby, prevent their re-oxidisation. The alkylation reaction was immediately covered in aluminium foil, placed in a cupboard and allowed to proceed in the dark for 30 minutes.

Both native TIL'E' and reduced and alkylated TIL'E' were blotted onto a nitrocellulose membrane in increasing amounts of 56.25 ng, 112.5 ng, 225 ng and 450 ng. The

membrane was allowed to dry at room temperature for one hour. Following washing with TBS buffer the membrane was incubated with blocking buffer (5% non-fat milk in TBS-T) for one hour at room temperature. The membrane was then washed three times; twice with TBS-T buffer and once with TBS buffer, and subsequently incubated with mAb-418 (diluted 1 in 1000) in blocking buffer for two hours at room temperature. Following washing, twice in TBS-T and once in TBS buffer, the membrane was then incubated with secondary HRP-conjugated anti-mouse IgG antibody (diluted 1 in 1000) in blocking buffer for one hour at room temperature. The membrane was washed three times (twice in TBS-T, then once in TBS buffer) prior to ECL chemiluminescence detection to reveal binding of mAb-418 to TIL'E'.

4.11. NMR experimental setup

NMR experiments were recorded by Dr D. Flemming Hansen and Dr John Kirkpatrick, at 25°C (298 K) on a Varian INOVA 600 MHz spectrometer and Bruker Avance III 500 MHz and 700 MHz spectrometers. The 600 MHz and 700 MHz spectrometers were equipped with cryogenic inverse triple-resonance probes, and the 500 MHz spectrometer with a room temperature inverse triple-resonance probe. NMR data were processed with NMRPipe[405]. Spectral analysis, peak picking and assignments were performed using the programme Sparky (Goddard TD and Kneller DG, SPARKY 3, University of California, San Francisco).

4.12. TIL'E' samples for NMR studies

^{15}N- and $^{13}C^{15}N$-labelled TIL'E' samples were prepared for NMR spectroscopy at concentrations of 0.5-1 mM in 20 mM sodium phosphate (Na_2HPO_4/NaH_2PO_4), 100 mM NaCl, 10% D_2O at pH 7.4. Samples used for backbone RDC measurements were prepared by adding *Pf*1 bacteriophage (ASLA Biotech) at a concentration of 6 mg/mL or $C_{12}E_5$/*n*-hexanol (Sigma-Aldrich) at 3% w/w $C_{12}E_5$, r = 0.94 (were r is the molar ratio of $C_{12}E_5$ to hexanol).

4.13. Chemical shift assignment

Backbone chemical shift assignments were obtained using standard triple-resonance experiments recorded at 600 MHz, including HNCO, HNCA, HN(CO)CA, HN(CA)CO and HNCACB spectra, supplemented with 15N-HSQC-NOESY-HSQC and HN(CACO)NH spectra. Side chain assignments were obtained from CC-TOCSY (total correlation) type experiments recorded at 600 MHz, including CC(CO)NH-TOCSY, HCC(CO)NH-TOCSY and HCCH-TOCSY. Each NH strip of the CC(CO)NH-TOCSY spectrum, will contain peaks for all carbon nuclei in the side chain of residue i-1, whereas, each NH strip of the HCC(CO)NH-TOCSY spectrum will contain peaks for all hydrogen nuclei in the side chain of residue i-1. In the HCCH-TOCSY spectrum, each CH strip will contain peaks for all the hydrogen atoms attached to all other carbon atoms in the same side chain. The prochiral methyl groups of valine and leucine were stereospecifically assigned according to the method of Neri *et al*[349]. The TIL'E' assigned chemical shift list is found in Appendix 5.2.

4.14. Backbone dihedral angles and RCI-S^2 order parameters

Backbone dihedral angle constraints, ϕ and ψ, were calculated from the backbone chemical shifts (^{15}N, $^{1}H^{N}$, ^{13}CO, $^{13}C\alpha$ and $^{13}C\beta$) using the programme TALOS+[345]. Dihedral angles that were classified as dynamic or bad by the TALOS+ algorithm[345] were excluded; 108 angles were used as restraints in the TIL'E' structure calculations. The TALOS+ programme[345] also calculates the RCI-derived order parameter S^2 using the backbone chemical shifts; this calculation is based on the method of Berjanski and Wishart[344]; a table of predicted RCI-S^2 values for TIL'E' is given in Appendix 5.4.

4.15. NOESY experiments

^{1}H-^{1}H distance restraints for TIL'E' structure calculations were extracted from two nuclear Overhauser enhancement spectra (NOESY); ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC spectra, recorded at 600 MHz (140 ms mixing time) and 700 MHz (120 ms mixing time), respectively. In the ^{15}N-NOESY- HSQC each strip contains NOEs from one NH group to all other hydrogen atoms in close proximity, whereas, each strip of the ^{13}C-

NOESY-HSQC spectrum contains NOEs from one CH group to all other hydrogens nearby. NOEs were manually assigned, in all three dimensions, and in total 954 unique inter-residue distance restraints were obtained (Table 4.9). The intensities of the NOEs were used as input for the programme CYANA 3.0 in subsequent structure calculations. In addition, 1540 intra-residue NOEs were assigned but where not used in the structure calculation, as they provide very limited structural information.

Table 4.9: Summary of NOE distance restraints.

Unique ^{1}H-^{1}H NOEs	Number
Inter-residue	954
Sequential(1)	383
Short-range(2-3)	121
Medium-range(4-5)	97
Long-range(>5)	353

4.16. Residual dipolar couplings

Vector orientations were calculated from experimental residual dipolar couplings (RDCs) obtained in two nearly orthogonal alignment media, *Pf*1 phage and $C_{12}E_5$/*n*-hexanol. $^{1}H^{N}_{i}$-$^{15}N_i$ RDCs were measured at 500 MHz in both alignment media using a 2D IPAP-^{15}N-HSQC pulse sequence. $^{15}N_i$-$^{13}CO_{i-1}$, $^{1}H^{N}_{i}$-$^{13}CO_{i-1}$, $^{13}CO_i$-$^{13}C\alpha_i$, $^{1}H^{N}_{i}$-$^{13}C\alpha_{i-1}$, $^{13}C\alpha_i$-$^{13}C\beta_i$ and $^{13}C\alpha_i$-$^{1}H\alpha_i$ RDCs were measured at 500 MHz in the $C_{12}E_5$/*n*-hexanol alignment medium only. $^{15}N_i$-$^{13}CO_{i-1}$ and $^{1}H^{N}_{i}$-$^{13}CO_{i-1}$ RDCs were measured using a 2D IPAP-E.COSY-^{15}N-HSQC experiment. $^{13}CO_i$-$^{13}C\alpha_i$ and $^{1}H^{N}_{i}$-$^{13}C\alpha_{i-1}$ RDCs were measured using a 3D IPAP-HNCO(J-CA) sequence; $^{13}C\alpha_i$-$^{13}C\beta_i$ RDCs were measured using a 3D in-phase HN(CO)CA(J-CB) sequence; $^{13}C\alpha_i$-$^{1}H\alpha_i$ RDCs were measured using a 3D IPAP-(HA)CA(CO)NH sequence. A total of 361 RDC restraints were obtained. VWF TIL'E' residues undergoing significant internal dynamics, as indicated by a ^{1}[H]-^{15}N NOE value of 0.6 or below or where the ^{1}[H]-^{15}N NOE could not be obtained, were excluded from the single structure calculations. However, all RDC measurements were included in the ensemble refinement calculation, which simultaneously represents the TIL'E' structure and its intrinsic picosecond to microsecond dynamics. Back-calculation of RDCs from

an RDC-refined TIL'E' NMR structure was undertaken with the programme Module 1.0[363].

4.17. TIL'E' structure calculation with CYANA 3.0

Lists that include assigned peak positions and peak intensities of the ^{13}C-edited and ^{15}N-edited NOESY spectra and assigned chemical shifts were exported from Sparky (Goddard TD and Kneller DG, SPARKY 3, University of California, San Francisco) and used as input files for CYANA 3.0[364]. Backbone dihedral angles obtained from TALOS+[345] were included in the standard manner with the command *read aco*. The eight TIL'E' disulfide bonds (Cys767-Cys808, Cys776- Cys804, Cys788- Cys799, Cys792-Cys827, Cys810-Cys821, Cys829-Cys851, Cys846-Cys863 and Cys849-Cys858) were explicitly implemented as constraints during the calculation using the *ssbond* command in CYANA 3.0. The stereospecifically assigned Val and Leu methyl groups were entered into the initialisation file with the command *atom stereo* for Val772, Val775, Val805, Val815, Val822, Val842, Val850, Val862 and Leu774, Leu781, Leu786, Leu797, Leu809, Leu824.

Vector orientation restraints (RDCs) were included with the suggested weight of (W_{RDC}) 0.01 and initial starting parameters for the alignment tensor, D_{max} = 15.7 Hz, R = 0.45 ($C_{12}E_5$/*n*-hexanol) and D_{max} = -15.7 Hz, R=0.45 (*Pf*1 phage) were estimated from histograms of the experimentally obtained RDCs. However, both the orientation and the size of the alignment tensors were optimised during the structure calculation and it was found that the specific starting values of the alignment tensors did not change the final family of structures.

The command *peakcheck* was used to check the NOESY spectra peak lists against the assigned chemical shift list for errors in assignment. Automatic distance calibration implemented in CYANA 3.0[364], converts NOE peak intensities into distance constraints. Upper distance limits are applied using an average distance limit for automatic calibration of 5 Å and the command *calibration*. The *distance modify* command was used to remove redundant and irrelevant restraints such as reciprocal NOEs, NOEs for which there exists no conformation that would violate it, and fixed distance NOEs such

as between geminal hydrogen atoms. The *distances check* command was used to assess how well supported an NOE is by surrounding NOEs. Finally, a total of 100 structures were calculated using 10000 steps of torsion angle molecular dynamics and the 10 structures with lowest energy, and no NOE violations greater than 0.5 Å were selected to represent the solution structure of VWF TIL'E'.

4.18. Heteronuclear ^{1}[H]-^{15}N NOE and CPMG experiments

Differential flexibility on a picosecond to nanosecond timescale for the TIL'E' molecule was quantitatively assessed by backbone heteronuclear ^{1}[H]-^{15}N NOE experiments and on the microsecond to millisecond timescale by constant time TROSY-based CPMG relaxation dispersion experiments, which were recorded at two field strengths (500 MHz and 700 MHz). For both experiments peak intensities were extracted by the programme FuDA[406]. For the CPMG experiments, peak intensities for the different values of ν_{CPMG} and I_{CPMG} were converted to effective transverse relaxation rates (R_2^{eff}) according to:

$$R_2^{eff}(\mathrm{CPMG}) = \frac{1}{\Delta} \ln \frac{I_{ref}}{I_{CPMG}(\nu_{CPMG})}$$

An approximation of the exchange contribution, R_{ex}, was subsequently calculated by :

$$R_{ex} \sim R_2^{eff}(CPMG_{min}) - R_2^{eff}(CPMG_{max})$$

A table of all TIL'E' RCI-S^2, ^{1}[H]-^{15}N NOE and R_{ex} values are given in Appendix 5.4.

4.19. Ensemble refinement calculation

The ensemble refinement calculations were performed by Dr D. Flemming Hansen using the programme Xplor-NIH. RDCs and RCI-derived S^2 order parameters from backbone chemical shifts (TALOS+)[345] were included as dynamical restraints. All experimentally determined RDCs were used. A twelve-member ensemble was calculated and two full sets of RDCs (^{15}N-CO and ^{13}CO-^{13}Cα) were initially excluded and used for cross-validation along with ^{13}C methyl chemical shifts.

5. Appendix

5.1. Intron-exon junctions in proVWF

Domain	Exon (length bp)	5' Splice donor	3' Splice acceptor	Splice junction phase
VWFD-C8-TIL-E				
VWD1	3 (165)	ATT.G Ile73.Gly74	GG.GAC Gly74.Asp75	1
VWD1	4 (103)	CAA.AG Gln107.Arg108	A.GTC Arg108.Val109	2
VWD1	5 (209)	GAA.G Glu177.Gly178	GG.ACC Gly178.Thr179	1
VWD1	6 (125)	TAT.GA Tyr185.Asp186	C.TTT Asp186.Phe187	2
C81 to TIL1 linker	7 (217)	TGC.A Cys291.Ser292	GC.CCA Ser292.Pro293	1
TIL1	8 (123)	CCT.G Pro332.Glu333	AG.GGA Glu333.Gly334	1
E1	9 (112)	ACC.TG Thr369.Cys370	C.ATT Cys370.Ile371	2
E1 to VWD2 linker	10 (47)	CCA.G Pro385.Gly386	GG.GAG Gly386.Glu387	1
VWD2	11 (137)	CAG Gln431	TGT Cys432	0
VWD2	12 (139)	AAA.G Lys477.Gly478	GT.GAC Gly478.Asp479	1
VWD2	13 (101)	AAG Lys511	CTG Leu512	0
C82	14 (196)	ATG.A Met576.Thr577	CC.AGG Thr577.Arg578	1
C82-TIL2 Linker	15 (216)	TGT.G Cys648.Glu649	AG.CTG Glu649.Leu650	1
E2	16 (241)	ATG.TG Met728.Cys729	C.TAC Cys729.Tyr728	2
E2 to TIL' linker	17 (95)	CGC.A Arg760.Ser761	GC.AAA Ser761.Lys762	1
TIL'	18 (161)	ATG Met814	GTC Val815	0
E'	19 (104)	ACT.TG Thr848.Cys849	T.GTC Cys849.Val850	2
VWD3	20 (139)	CAG Gln895	GAT Asp896	0
VWD3	21 (135)	GAG Glu940	GTG Val941	0
VWD3	22 (147)	CAG Gln989	GAG Glu990	0
VWD3 to C8 linker	23 (141)	AAA Lys1036	GTG Val1037	0
C83	24 (114)	CTG Leu1074	GTG Val1075	0
C83 to TIL3 Linker	25 (157)	TGC.C Cys1126.Pro1127	CC.CAG Pro1127.Gln1128	1

TIL3	26 (159)	CCA.G Pro1179.Gly1180	GG.AAA Gly1180.Lys1181	1
E3	27 (136)	ATT.TG Ile1224.Cys1225	C.CAC Cys1225.His1226	2
D4N	33 (44)	AGG Arg1888	CCC Pro1889	0
D4N	34 (178)	CCC.T Pro1947.Cys1948	GC.GTG Cys1948.Val1949	1
VWD4	35 (221)	GAG Glu2021	GTG Val2022	0
VWD4	36 (193)	TGT.G Cys2085.Gly2086	GG.ATC Gly2086.Ile2087	1
C84 to TIL4 Linker	37 (342)	TGT.G Cys2199.Ala2200	CT.ATG Ala2200.Met2201	1
VWC				
VWC1	38 (200)	CAG Gln2266	TTC Phe2267	0
VWC1	39 (103)	AAA.G Lys2300.Ala2301	CT.CCC Ala2301.Pro2302	1
VWC1 to VWC2 linker	40 (75)	TGT.G Cys2325.Val2326	TG.TGT Val2326.Cys2327	1
VWC2	41 (105)	TGC.G Cys2360.Ala2361	CC.TGC Ala2361.Cys2362	1
VWC2 to VWC3 linker	42 (206)	AAG Lys2429	GTG Val2430	0
VWC3	43 (150)	TCG Ser2479	GGC Gly2480	0
VWC4	44 (111)	AGT Ser2516	GTC Val2517	0
VWC4 to VWC5 linker	45 (181)	TGT.G Cys2576.Glu2577	AG.CGC Glu2577.Arg2578	1
VWC5	46 (41)	GGG Gly2590	CCC Pro2591	0
VWC5	47 (117)	CTG Leu2629	GGT Gly2630	0
VWC6	48 (99)	AAG Lys2662	CTG Arg2663	0
VWC6	49 (129)	GGA Gly2705	GGT Gly2706	0
VWC6	50 (40)	ACA.T Thr2718.Cys2719	GT.GAG Cys2719.Glu2720	1

5.2. VWF TIL'E' chemical shift assignment

Chemical shifts are given in ppm; no chemical shift indicates unassigned or unobserved resonances.

Residue	H^N	Hα	Hβ	Hγ	Hδ	Hε	Cα	Cβ	Cγ	Cδ	Cε	N	Nδ	Nε
S766														
C767	8.184											123.547		
R768														
P769		4.565	2.323,1.922		3.791,3.555		62.078	30.636		50.454				
P770		4.564	2.305,2.152	2.008,1.437	3.524,3.480		64.007	33.632	24.545	49.539				
M771	8.759	4.346	2.056,1.782	2.725,2.271		1.928	57.932	33.859	34.256		18.044	123.176		
V772	8.855	4.301	1.974	0.887,0.825			59.732	35.145	21.357,19.573			116.266		
K773	8.448	4.371	1.770,1.605	1.242,1.210			56.715	33.170	25.207			125.383		
L774	9.145	4.549	1.919,1.498	1.592	0.865,0.901		54.112	43.288	28.227	24.974,25.383		132.136		
V775	8.578	4.149	1.931	0.860,0.864			61.132	32.365	20.987,20.981			127.590		
C776	8.643	4.291	2.962,2.691				56.014	40.082				129.201		
P777		4.497	2.290,2.100	2.154,2.062	3.801,3.613		62.350	32.151	27.459	50.442				
A778	8.553	4.120	1.376				53.417	18.882				124.507		
D779	8.116	4.505	2.748,2.635				53.654	40.387				114.322		
N780	7.669	4.656	2.809,2.813		7.623,6.892		53.230	38.076				115.834	111.505	
L781		4.287	1.659,1.596		0.918,0.847		55.807	41.758		25.022,23.143				
R782	7.897	4.303	1.946,1.736	1.642,1.540	3.180,3.182		55.632	29.732	27.129	43.128		117.867		
A783	7.345	4.215	1.349				52.261	20.466				123.746		
E784	8.061	4.542	2.273,1.885	2.131,1.710			56.544					119.157		
G785														
L786		4.315	1.666,1.576	1.600	0.895,0.835		55.313	42.152	26.986	25.081,23.503				
E787	8.165											121.770		
C788														
T789			4.085	1.224				70.366	21.424					
K790	8.762	5.059	1.863,1.845	1.655,1.621		2.925,2.928	54.874	34.598	24.691		42.477	128.269		
T791	9.410	5.450	4.953	1.217			58.958	72.724	21.387			116.019		
C792	8.587	4.376	3.181,3.034				57.480	36.316				115.821		
Q793	8.303	4.127	1.945,1.863	2.155,2.162		7.329,6.751	58.102	29.341	34.424			115.926		111.938

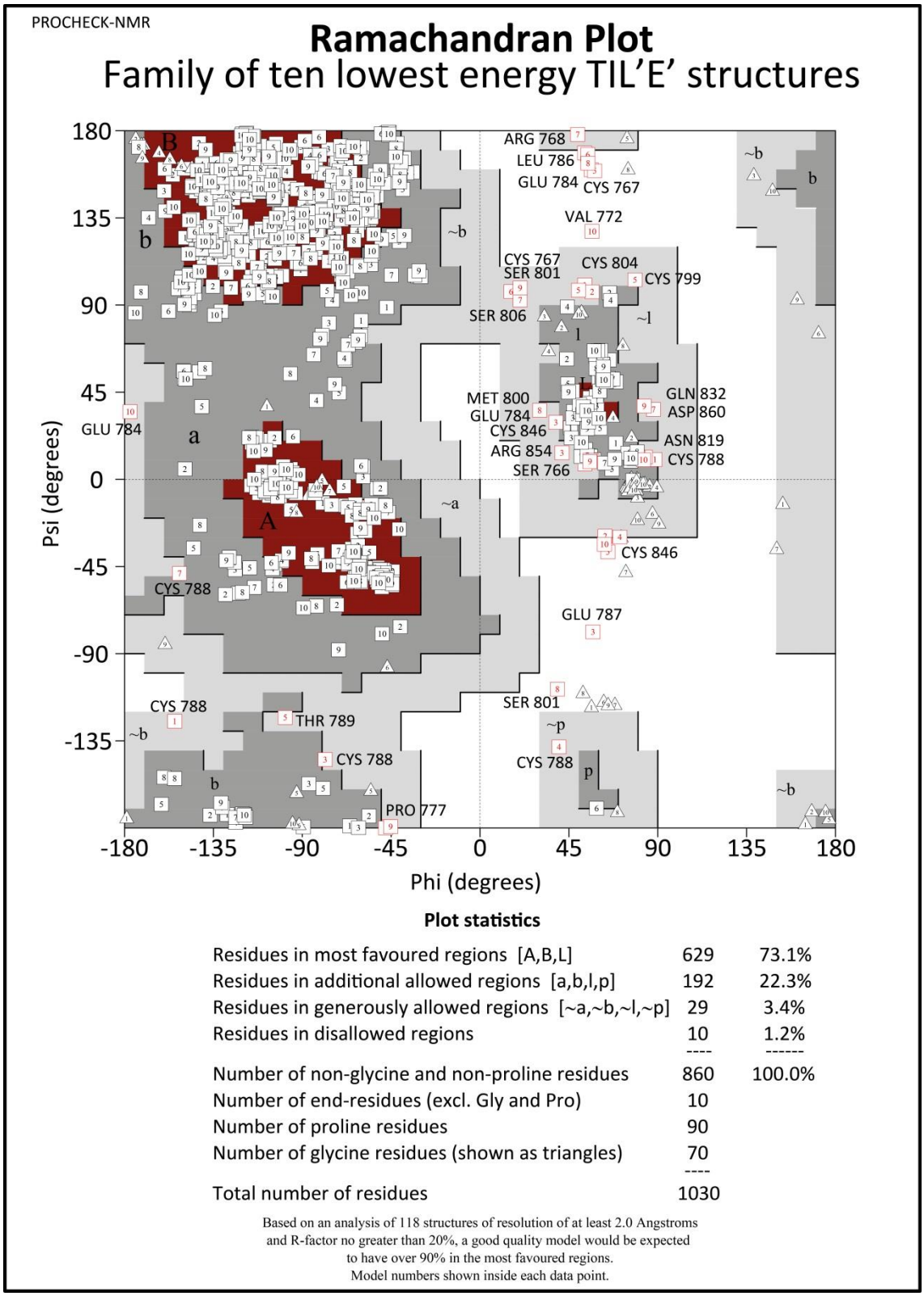
Residue	H^N	Hα	Hβ	Hγ	Hδ	Hε	Cα	Cβ	Cγ	Cδ	Cε	N	Nδ	Nε
N794	7.925	4.760	3.316,2.628		7.396,7.010		51.927	39.325				117.107	108.104	
Y795	7.310	3.885	3.384,2.739		6.952	6.702	63.779	39.078		133.344	118.227	120.190		
D796	8.291	4.507	2.793,2.472				54.482	41.078				113.579		
L797	7.787	4.406	1.742,1.490	1.566	0.912,0.813		53.959	42.569	26.865	25.301,22.659		120.878		
E798	8.490	4.120	1.909,1.863	2.241,2.168			56.295	30.252	26.298			122.509		
C799														
M800		4.588	2.008,1.925	2.498,2.414			54.180	31.754	32.049					
S801	8.374	4.288	3.924,3.698				58.207	63.324				118.194		
M802		4.677		2.687,2.491			54.482		32.366					
G803	8.035	4.081,3.601					45.423					111.120		
C804	8.463	4.556	3.105,2.843					38.600				126.270		
V805		4.218	2.007	1.055,1.029			62.334	34.050	21.145,21.330					
S806	8.484	4.528	4.186,3.848				59.211	64.191				126.084		
G807	8.516	4.320,3.916					45.758					107.758		
C808	8.554	5.227	2.825,2.798				54.413	41.276				120.532		
L809	9.410	4.686	1.737,1.600	1.748	0.883,0.834		53.774	45.572	27.044	25.905,24.858		125.794		
C810	9.188	4.522	2.271,2.144				54.090	36.633				119.162		
P811														
P812														
G813		4.328,3.662					44.936							
M814	8.297	5.043	1.957,1.755	2.447,2.050		1.933	54.425	37.090	33.534		18.403	118.736		
V815	9.126	4.576	1.825	0.734,0.571			58.879	35.361	23.542,19.578			113.319		
R816	8.490	4.980	1.821,1.776	1.633,1.520	3.112,3.075		55.602	31.287	27.370	43.535		122.532		
H817	9.100	4.858	3.006,2.840		6.896	7.793	56.971	34.758		117.027	138.910	126.203		
E818	8.953	3.563	1.915,1.893	1.714,1.587			57.676	26.921	35.712			125.530		
N819	8.687	4.407	2.986,2.866		7.569,6.904		53.973	38.074				114.182	113.597	
R820	8.027	4.750	1.893,1.832	1.750,1.572	3.190,3.192		54.554	33.324	26.545	43.533		118.232		
C821	8.703	5.477	2.878,2.728				53.980	40.976				117.965		
V822	9.671	4.625	2.025	0.885,0.691			58.996	36.362	21.682,19.875			119.418		
A823	8.552	4.356	1.388				52.418	18.618				125.768		
L824	7.979	3.782	1.410,1.135	1.123	0.182,0.310		58.182	42.411	26.685	22.890,25.021		123.359		
E825	8.950	4.025	2.018,2.008	2.165,2.167			58.502	28.267	36.068			113.695		
R826	8.153	4.245	2.060,1.835	1.741,1.528	3.174,3.171		54.554	29.158	26.950	42.526		118.469		

Residue	H^N	Hα	Hβ	Hγ	Hδ	Hε	Cα	Cβ	Cγ	Cδ	Cε	N	Nδ	Nε
C827	7.601	4.224	3.019,2.744				55.125	37.198				119.490		
P828		4.437	1.970,1.496	1.682,1.426	3.226,3.170		62.458	33.668	26.386	49.726				
C829	8.424	4.328	1.955,1.536				56.386	42.435				112.961		
F830	8.799	5.831	3.080,2.956		7.109	7.239	57.1245	42.551		132.054	131.598	118.787		
H831	9.441	4.968	2.988,2.868		6.682	7.482	57.755	33.828		118.872	138.664	123.887		
Q832		3.733	2.087,1.650	1.746,1.334		7.273,6.785	56.512	26.732	33.404					111.289
G833	8.665	4.206,3.693					45.486					104.219		
K834	7.951	4.670	1.901,1.832	1.430,1.208	1.677,1.680	2.918,2.917	54.931	35.124	25.305	28.949	42.349	122.208		
E835	8.216	4.332	1.828,1.704	1.942,1.841			55.614	31.797	36.920			122.599		
Y836	8.937	4.431	2.809,2.502		7.227	6.990	57.231	41.874		133.347	118.515	124.173		
A837	8.676	4.357	1.295				49.933	17.807				124.606		
P838		3.854	2.260,1.774	1.979,1.937	3.478,3.389		63.884	31.167	28.326	50.032				
G839	9.487	4.369,3.407					44.966					113.317		
E840	8.254	4.486	2.553,2.416	2.562,2.337			57.610	31.261	38.679			120.804		
T841	8.132	5.752	3.957	1.096			59.984	72.217	21.464			110.916		
V842	8.459	4.204	1.081	-0.411, -0.446			59.688	35.482	20.212,18.397			117.320		
K843	8.017	4.811	1.667,1.509	1.249,1.221	1.531,1.530	2.868,2.808	54.855	33.297	25.529	29.130	41.898			
I844	8.039	3.995	1.478	0.703 0.839,0.246	0.076		60.858	38.347	26.765,16.839	12.200		125.653		
G845		3.984,3.662					47.357							
C846	8.522	4.929	3.191,3.189				58.558	37.971				122.743		
N847	8.165	4.923	2.724,2.451		8.262,7.901		53.858	41.629				121.769	118.942	
T848	8.250	4.613	3.936	1.094			62.341	70.460	21.606			115.576		
C849	9.054	5.544	2.732,2.595				54.867	50.459				124.212		
V850	8.888	5.046	1.971	0.940,0.923			59.590	35.195	20.202,21.444			122.162		
C851	8.382	4.170	2.585,1.798				54.727	36.915				128.312		
R852	9.051	4.309	1.654,1.650	1.538,1.470	3.170,3.115		56.304	32.883	27.026	43.304		132.067		
D853	9.328	4.204	2.732,2.565				55.745	39.142				128.437		
R854	8.607	3.274	2.065,1.915	1.443,1.385	3.146,2.981		58.401	28.561	28.766	43.223		108.225		
K855	7.712	4.524	1.756,1.701	1.411,1.367	1.725,1.671	3.024,3.028	53.875	36.106	24.041	29.110	42.201	121.587		
W856	8.366	4.812	3.024,2.881		7.353	9.618,6.723	55.886	30.257		127.360	119.588	120.035		126.126
N857	9.268	5.068	2.917,2.776		7.093,6.919		52.832	39.532				125.356	113.428	

Residue	H^N	Hα	Hβ	Hγ	Hδ	Hε	Cα	Cβ	Cγ	Cδ	Cε	N	Nδ	Nε
C858		5.712	3.016,2.977				55.808	47.903						
T859	8.221	4.183	4.012	1.358			62.852	70.176	21.732			112.700		
D860	8.038	4.649	2.554,2.496				52.707	40.809				120.286		
H861	8.654	4.655	3.041,2.938		7.044	7.854	55.554	30.875		120.696	138.890	120.257		
V862	8.687	3.982	2.008	0.944,0.999			62.929	31.843	21.349,21.206			123.478		
C863	8.700	4.844	3.532,2.831				52.964	37.591				125.981		
D864	8.041	4.345	2.618,2.534				55.939	42.176				127.664		

The assigned atoms, Cζ (129.002) and Hζ (7.131) of Phe830 and $C\eta_2$ (123.627), $C\zeta_2$ (113.591), $C\zeta_3$ (123.071), $H\eta_2$ (6.092), $H\zeta_2$ (6.552) and $H\zeta_3$ (6.768) of Trp856 are not included in the TIL'E' shift table due to space constraints.

5.3. Ramachandran plot of ten lowest energy TIL'E' NMR structures



Plot statistics

Residues in most favoured regions [A,B,L]	629	73.1%
Residues in additional allowed regions [a,b,l,p]	192	22.3%
Residues in generously allowed regions [~a,~b,~l,~p]	29	3.4%
Residues in disallowed regions	10	1.2%
Number of non-glycine and non-proline residues	860	100.0%
Number of end-residues (excl. Gly and Pro)	10	
Number of proline residues	90	
Number of glycine residues (shown as triangles)	70	
Total number of residues	1030	

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.

5.4. Picosecond to millisecond dynamics of TIL'E'

Dynamic residues, [^{1}H]-^{15}N NOEs and RCI-S^2 less than 0.6 and R_{ex} greater than 1.5 s^{-1} are coloured red.

Residue	Secondary structure	[^{1}H]-^{15}N NOEs (ps-ns)	RCI-S^2 (ps-μs)	R_{ex} (s^{-1}; μs-ms)
S766				
C767			0.788	
R768				
P769			0.734	
P770			0.754	
M771		0.74±0.02	0.764	**3.62 ± 0.07**
V772	**TIL' β1**	**0.60 ± 0.04**	0.793	**8.40 ± 0.50**
K773		0.70 ± 0.02	0.784	**2.17 ± 0.20**
L774		0.69 ± 0.08	0.813	
V775		0.63 ± 0.02	0.815	**2.42 ± 0.20**
C776		0.66 ± 0.04	0.817	**16.23 ± 1.07**
P777			0.766	
A778		**0.46 ± 0.07**	0.730	**2.75 ± 1.49**
D779		**0.58 ± 0.02**	0.704	**13.64 ± 0.65**
N780			0.677	**7.54 ± 0.85**
L781			0.642	
R782		**0.42 ± 0.05**	**0.601**	**14.00 ± 0.15**
A783			**0.551**	
E784		**0.52 ± 0.03**	**0.524**	**5.51 ± 0.04**
G785				
L786			**0.528**	
E787			**0.528**	
C788				
T789			0.844	
K790	**Upper portion of TIL' β1-to-β2 loop**	0.76 ± 0.03	0.862	**6.57 ± 0.12**
T791		0.65 ± 0.02	0.884	**8.50 ± 0.63**
C792		0.72 ± 0.01	0.875	0.67 ± 0.26
Q793		0.64 ± 0.03	0.860	**3.02 ± 0.38**
N794		0.69 ± 0.02	0.853	**1.56 ± 0.07**
Y795		0.67 ± 0.01	0.865	**2.89 ± 0.09**
D796		0.75 ± 0.02	0.852	**2.76 ± 0.05**
L797		0.72 ± 0.01	0.819	**1.84 ± 0.10**
E798			0.794	
C799				
M800			0.671	
S801		**0.42 ± 0.02**	0.658	**13.31 ± 0.38**
M802			0.674	
G803		**0.55 ± 0.03**	0.704	**19.13 ± 0.27**
C804				
V805			0.699	
S806	**TIL' β2**	0.70 ± 0.07	0.742	**16.10 ± 0.42**
G807		0.72 ± 0.03	0.772	**14.74 ± 0.19**
C808		0.67 ± 0.03	0.867	-0.28 ± 0.18
L809		0.75 ± 0.04	0.862	**6.22 ± 0.65**
C810			0.858	
P811				
P812				
G813			0.786	
M814	**TIL' β3**	0.68 ± 0.01	0.818	0.33 ± 0.04

Residue	Secondary structure	[^{1}H]-^{15}N NOEs (ps-ns)	RCI-S^2 (ps-μs)	R_{ex} ($s^{-1;}$ μs-ms)
V815		0.72 ± 0.02	0.885	**5.12 ± 0.28**
R816			0.886	
H817		0.78 ± 0.02	0.879	**5.92 ± 0.35**
E818		0.68 ± 0.02	0.871	-0.60 ± 0.22
N819			0.881	
R820		0.68 ± 0.01	0.897	0.58 ± 0.09
C821	TIL' β4	0.61 ± 0.02	0.907	0.42 ± 0.10
V822		0.75 ± 0.02	0.884	0.49 ± 0.20
A823		0.77 ± 0.02	0.859	1.29 ±0.01
L824		0.77 ± 0.04	0.849	-0.21 ± 0.15
E825	3_{10}-helix	0.74 ± 0.01	0.858	0.26 ± 0.11
R826		0.70 ± 0.03	0.841	0.70 ± 0.11
C827		0.76 ± 0.01	0.835	0.68 ± 0.06
P828			0.835	
C829		0.73 ± 0.02	0.868	0.13 ± 0.10
F830	E' β1	0.71 ± 0.01	0.876	0.57 ± 0.01
H831		0.72 ± 0.01	0.851	0.17 ± 0.12
Q832			0.815	
G833		0.73 ± 0.03	0.753	-0.08 ± 0.15
K834		0.69 ± 0.01	0.722	-0.16 ± 0.06
E835	E' β2	0.61 ± 0.01	0.679	-0.10 ± 0.13
Y836		0.68 ± 0.02	0.688	0.01 ± 0.00
A837		**0.56 ± 0.01**	0.696	-0.40 ±0.03
P838			0.713	
G839		0.70 ± 0.01	0.743	-0.20 ± 0.01
E840		0.73 ± 0.02	0.784	-0.01 ± 0.06
T841	E' β3	0.72 ± 0.01	0.855	0.07 ± 0.09
V842		0.70 ± 0.01	0.864	0.50 ± 0.08
K843		0.66 ± 0.02	0.847	-0.17 ± 0.03
I844		0.63 ± 0.01	0.821	-0.39 ± 0.02
G845			0.806	
C846			0.778	
N847			0.796	
T848		0.65 ± 0.01	0.834	-0.12 ± 0.02
C849	E' β4	0.66 ± 0.01	0.907	-0.48 ± 0.05
V850		0.70 ± 0.01	0.919	-0.10 ± 0.09
C851		0.65 ± 0.01	0.915	-0.91 ± 0.10
R852		0.74 ± 0.01	0.905	
D853		0.65 ± 0.01	0.908	-1.08 ± 0.01
R854		0.66 ± 0.01	0.901	0.06 ± 0.11
K855		0.70 ± 0.01	0.859	0.28 ± 0.05
W856	E' β5	0.63 ± 0.01	0.853	0.98 ± 0.04
N857		0.66 ± 0.01	0.852	-0.55 ± 0.09
C858			0.879	
T859		0.67 ± 0.01	0.782	0.32 ± 0.06
D860		**0.58 ± 0.01**	0.710	0.36 ± 0.01
H861		**0.58 ± 0.01**	**0.607**	-0.07 ± 0.06
V862		**0.51 ± 0.01**	**0.561**	-0.30 ± 0.02
C863		**0.40 ± 0.01**	**0.514**	-0.96 ± 0.14
D864		**-0.15 ± 0.01**	**0.504**	-0.62 ± 0.03

Data from both experiments were not obtainable for all assigned residues owing to some peaks being very weak or overlapped and consequently excluded from further analysis.

6. References

1. Starr D. Blood: an epic history of medicine and commerce. 2000.

2. Ruggeri Z. Platelet adhesion under flow. *Microcirculation*. 2009;16(1):58–83.

3. Macfarlane R. An enzyme cascade in the blood clotting mechanism, and its function as a biochemical amplifier. *Nature*. 1964;202:498–499.

4. Davie EW, Ratnoff OD. Waterfall sequence for intrinsic blood clotting. *Science*. 1964;145(3638):1310–1312.

5. Hoffman M. A cell-based model of coagulation and the role of factor Vlla. *Blood Reviews*. 2003;17:51–55.

6. Monroe DM, Hoffman M. What does it take to make the perfect clot? *Arteriosclerosis thrombosis and vascular biology*. 2006;26(1):41–48.

7. Versteeg HH, Heemskerk JWM, Levi M, Reitsma PH. New fundamentals in hemostasis. *Physiological reviews*. 2013;93(1):327–58.

8. Ruggeri ZM. The role of von willebrand factor in thrombus formation. *Thrombosis research*. 2007;120(Suppl 1):S5–S9.

9. Terraube V, O'Donnell JS, Jenkins PV. Factor VIII and von willebrand factor interaction: biological, clinical and therapeutic importance. *Haemophilia the official journal of the World Federation of Hemophilia*. 2010;16(1):3–13.

10. Goodeve AC. The genetic basis of von willebrand disease. *Blood Reviews*. 2010;24(3):123–134.

11. Cumming AM, Keeney S, Jenkins PV, Nash MJ, O'Donnell JS. Clinical utility gene card for: von willebrand disease. *European journal of human genetics*. 2011;19(5):1–4.

12. Nichols WL, Hultin MB, James a H, et al. Von willebrand disease (VWD): evidence-based diagnosis and management guidelines, the national heart, lung, and blood institute (NHLBI) expert panel report (USA). *Haemophilia*. 2008;14(2):171–232.

13. Federici A, Castaman G, Thompson A, Berntorps E. Von willebrand ' s disease : clinical management. *Haemophilia*. 2006;12(Suppl. 3):152–158.

14. Lynch D, Zimmerman T, Collins C, et al. Molecular cloning of cDNA for human von willebrand factor: authentication by a new method. *Cell*. 1985;41:49–56.

15. Sadler JE, Shelton-Inloes BB, Sorace JM, et al. Cloning and characterization of two cDNA coding for human von willebrand factor. *Proceedings of the National Academy of Sciences*. 1985;82(19):6394–6398.

16. Verweij C, Vries C de, Distel B, et al. Construction of cDNA coding for human von willebrand factor using antibody probes for colony-screening and mapping of the chromosomal gene. *Nucleic Acids Research*. 1985;13(13):4699–4717.

17. Ginsburg D, Handin R, Bonthron D, et al. Human von willebrand factor (VWF): isolation of complementary dna (cdna) clones and chromosomal localization. *Science*. 1985;228(4706):1401–1406.

18. Kuwano A, Morimoto Y, Nagai T, et al. Precise chromosomal locations of the genes for dentatorubral-pallidoluysian atrophy (DRPLA), von willebrand factor (F8VWF) and parathyroid hormone-like hormone (PTHLH) in human chromosome 12p by deletion mapping. *Human Genetics*. 1996;97:95–98.

19. Mancuso DJ, Tuley EA, Westfield LA, et al. Structure of the gene for human von willebrand factor. *The Journal of Biological Chemistry*. 1989;264(33):19514–19527.

20. Shelton-Inloes BB, Chehab FF, Mannucci PM, Federici AB, Sadler JE. Gene deletions correlate with the development of alloantibodies in von willebrand disease. *Journal of Clinical Investigation*. 1987;79(5):1459–1465.

21. Patracchini P, Calzolari E, Aiello V, et al. Sublocalization of von willebrand factor pseudogene to 22q11. 22–q11. 23 by in situ hybridization in a 46,x,t(x; 22)(pter; q11. 21) translocation. *Human genetics*. 1989;83:264–266.

22. Mancuso DJ, Tuley E a, Westfield L a, et al. Human von willebrand factor gene and pseudogene: structural analysis and differentiation by polymerase chain reaction. *Biochemistry*. 1991;30:253–269.

23. Sadler J. Biochemistry and genetics of von willebrand factor. *Annual review of biochemistry*. 1998;67:395–424.

24. Eikenboom JC, Vink T, Briët E, Sixma JJ, Reitsma PH. Multiple substitutions in the von willebrand factor gene that mimic the pseudogene sequence. *Proceedings of the National Academy of Sciences*. 1994;91(6):2221–4.

25. Surdhar GK. Homozygous gene conversion in von willebrand factor gene as a cause of type 3 von willebrand disease and predisposition to inhibitor development. *Blood*. 2001;98(1):248–250.

26. Jaffe EA, Hoyer LW, Nachman RL. Synthesis of von willebrand factor by cultured human endothelial cells. *Proceedings of the National Academy of Sciences*. 1974;71(5):1906–1909.

27. Nachman R, Levine R, Jaffe EA. Synthesis of factor viii antigen by cultured guinea pig megakaryocytes. *Journal of Clinical Investigation*. 1977;60(4):914–921.

28. Bowie E, Jr LS, Fass D, et al. Transplantation of normal bone marrow into a pig with severe von willebrand's disease. *Journal of Clinical Investigation.* 1986;78:26–30.

29. Wagner D, Olmsted J, Marder V. Immunolocalization of von willebrand protein in Weibei-Palade bodies of human endothelial cells unlabeled peroxidase-anti-peroxidase staining. *The Journal of Cell Biology*. 1982;95(1):355–360.

30. Weibel R, Palade G. New cytoplasmic in arterial components morphology of the observed cytoplasmic. *The Journal of Cell Biology*. 1964;23(1):101–112.

31. 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 Biology International Reports*. 1984;8(10):

32. Wencel-Drake JD, Painter RG, Zimmerman TS, Ginsberg MH. Ultrastructural localization of human platelet thrombospondin, fibrinogen, fibronectin, and von Willebrand factor in frozen thin section. *Blood*. 1985;65(4):929–938.

33. 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):2791–2801.

34. 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 thrombosis and vascular biology*. 1987;7(3):287–291.

35. Ruggeri ZM. von Willebrand factor, platelets and endothelial cell interactions. *Journal of thrombosis and haemostasis*. 2003;1(7):1335–1342.

36. Senis Y, Richardson M, Tinlin S, Maurice D, Giles A. 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.

37. Aird W, Edelberg J, Weiler-Guettler H, et al. Vascular bed-specific expression of an endothelial cell gene is programmed by the tissue microenvironment. *The Journal of Cell Biology*. 1997;138(5):1117–1124.

38. Titani K, Kumar S, Takio K, et al. Amino acid sequence of human von willebrand factor. *Biochemistry*. 1986;25(11):3171–3184.

39. Verweij C, Diergaarde P, Hart M, Pannekoek H. Full-length von Willebrand factor (VWF) cDNA encodes a highly repetitive protein considerably larger than the mature vwf subunit. *The EMBO journal*. 1986;5(8):1839–1847.

40. Dang L, Purvis A, Huang R, Westfield L, Sadler J. Phylogenetic and functional analysis of histidine residues essential for ph-dependent multimerization of von willebrand factor. *Journal of Biological Chemistry.* 2011;286(29):25763–25769.

41. Fahey RC, Hunt JS, Windham GC. On the cysteine and cystine content of proteins: differences between intracellular and extracellular proteins. *Journal of Molecular Evolution*. 1977;10(2):155–160.

42. Marti T, Rösselet SJ, Titani K, Walsh KA. Identification of disulfide-bridged substructures within human von willebrand factor. *Biochemistry*. 1987;26(25):8099–8109.

43. Purvis AR, Gross J, Dang LT, et al. Two cys residues essential for von willebrand factor multimer assembly in the golgi. *Proceedings of the National Academy of Sciences*. 2007;104(40):15647–15652.

44. Choi H, Aboulfatova K, Pownall HJ, Cook R, Dong J. Shear-induced disulfide bond formation regulates adhesion activity of von Willebrand factor. *The Journal of Biological Chemistry*. 2007;282(49):35604–35611.

45. Li Y, Choi H, Zhou Z, et al. Covalent regulation of ULVWF string formation and elongation on endothelial cells under flow conditions. *Journal of thrombosis and haemostasis*. 2008;6(7):1135–1143.

46. Ganderton T, Berndt M, Chesterman C, Hogg P. Hypothesis for control of von Willebrand factor multimer size by intra-molecular thiol-disulphide exchange. *Journal of Thrombosis and Haemostasis*. 2007;5(1):204–6.

47. Shelton-Inloes BB, Titani K, Sadler JE. cDNA sequences for human von Willebrand factor reveal five types of repeated domains and five possible protein sequence polymorphisms. *Biochemistry*. 1986;25(11):3164–3171.

48. Bonthron DT, Handin RI, Kaufman RJ, et al. Structure of pre-pro-von Willebrand factor and its expression in heterologous cells. *Nature*. 1986;324(6094):270–3.

49. Shelton-Inloes BB, Broze G, Miletich JP, Sadler JE. Evolution of human von willebrand factor: cDNA seouence polynorphisms, repeated domain, and relationship to von willebrand antigen II. *Biochemical and Biophysical Research Communications*. 1987;144(2):657–665.

50. Zhou YF, Eng ET, Zhu J, et al. Sequence and structure relationships within von Willebrand factor. *Blood*. 2012;120(2):449–58.

51. Marcotte E, Pellegrini M, Yeates T, Eisenberg D. A census of protein repeats. *Journal of molecular biology*. 1999;293(1):151–160.

52. Andrade M, Perez-Iratxeta C, Ponting CP. Protein repeats: structures, functions, and evolution. *Journal of Structural Biology*. 2001;134(2-3):117–131.

53. Björklund AK, Ekman D, Elofsson A. Expansion of protein domain repeats. *PLoS computational biology*. 2006;2(8):e114.

54. Foster P, Fulcher C, Marti T, Titani K, Zimmerman T. A major factor VIII binding domain resides within the amino-terminal 272 amino acid residues of von Willebrand factor. *Journal of Biological Chemistry*. 1987;262(18):8443–8446.

55. Bockenstedt P, Greenberg JM, Handin RI. Structural basis of von Willebrand factor binding to platelet glycoprotein ib and collagen. effects of disulfide reduction and limited proteolysis of polymeric von willebrand factor. *Journal of Clinical Investigation*. 1986;77(3):743–749.

56. Fretto L, Fowler W, McCaslin D, Erickson H, McKee P. Substructure of human von Willebrand factor: proteolysis by V8 and characterisation of two functional domains. *The Journal of Biological Chemistry*. 1986;261(33):15679–15689.

57. Sixma JJ, Schiphorst ME, Verweij CL, Pannekoek H. Effect of deletion of the A1 domain of von Willebrand factor on its binding to heparin, collagen and platelets in the presence of ristocetin. *European journal of biochemistry*. 1991;196(2):369–375.

58. Fujimura Y, Titani K, Holland LZ, et al. A heparin-binding domain of human von Willebrand factor. *The Journal of Biological Chemistry*. 1987;262(4):1734–1739.

59. Mohri H, Yoshioka a, Zimmerman TS, Ruggeri ZM. Isolation of the von Willebrand factor domain interacting with platelet glycoprotein Ib, heparin, and collagen and characterization of its three distinct functional sites. *The Journal of biological chemistry*. 1989;264(29):17361–7.

60. Sobel M, Soler DF, Kermode JC, Harris RB. Localization and characterization of a heparin binding domain peptide of human von Willebrand factor. *JBiolChem*. 1992;267(13):8857–8862.

61. Fujimura Y, Titani K, Holland L, et al. Von Willebrand factor. a reduced and alkylated 52/48-kda fragment beginning at amino acid residue 449 contains the domain interacting with platelet glycoprotein Ib. *Journal of Biological Chemistry*. 1986;261(1):381–385.

62. Huizinga EG, Tsuji S, Romijn R a P, et al. Structures of glycoprotein Ibα and its complex with von Willebrand factor A1 domain. *Science*. 2002;297(August):1176–1179.

63. Dumas JJ, Kumar R, McDonagh T, 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. *The Journal of Biological Chemistry*. 2004;279(22):23327–23334.

64. Hoylaerts MF, Yamamoto H, Nuyts K, et al. von Willebrand factor binds to native collagen VI primarily via its A1 domain. *Biochemical Journal*. 1997;324(Part 1):185–191.

65. Mazzucato M, Spessotto P, Masotti A, et al. Identification of domains responsible for von Willebrand factor type VI collagen interaction mediating platelet adhesion under high flow. *The Journal of biological chemistry*. 1999;274(5):3033–41.

66. Romijn R A, Westein E, Bouma B, et al. Mapping the collagen-binding site in the von willebrand factor-A3 domain. *The Journal of Biological Chemistry*. 2003;278(17):15035–15039.

67. Nishida N, Sumikawa H, Sakakura M, et al. Collagen-binding mode of VWF-A3 domain determined by a transferred cross-saturation experiment. *Nature Structural Biology*. 2003;10(1):53–58.

68. Fujikawa K, Suzuki H, McMullen B, Chung D. Purification of human von Willebrand factor-cleaving protease and its identification as a new member of the metalloproteinase family. *Blood*. 2001;98(6):1662–1666.

69. Gao W, Anderson PJ, Sadler JE. Extensive contacts between ADAMTS13 exosites and von willebrand factor domain A2 contribute to substrate specificity. *Blood*. 2008;112(5):1713–1719.

70. Wu JJ, Fujikawa K, McMullen B a, Chung DW. Characterization of a core binding site for adamts-13 in the a2 domain of von willebrand factor. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(49):18470–18474.

71. Crawley JT, Groot R De, Xiang Y, Luken BM, Lane DA. Unravelling the scissile bond: how ADAMTS13 recognises and cleaves von Willebrand factor. *Blood*. 2011;118(12):3212–21.

72. Dent J, Berkowitz S, Ware J, Kasper C, Ruggeri Z. Identification of a cleavage site directing the immunochemical detection of molecular abnormalities in type iia von Willebrand factor. *Proceedings of the National Academy of Sciences*. 1990;87(8):6306–6310.

73. Ruggeri ZM, De Marco L, Gatti L, Bader R, Montgomery RR. Platelets have more than one binding site for von Willebrand factor. *Journal of Clinical Investigation*. 1983;72(1):1–12.

74. Berliner S, Niiya K, Roberts J, Houghten R, Ruggeri Z. Generation and characterization of peptide-specific antibodies that inhibit von willebrand factor binding to glycoprotein IIb-IIIa without interacting with other adhesive. *Journal of Biological* Chemistry. 1988;263(16):7500–7505.

75. 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):2637–2645.

76. Roberts DD, Williams SB, Gralnick HR, Ginsburg V. von Willebrand factor binds specifically to sulfated glycolipids. *The Journal of Biological Chemistry*. 1986;261(7):3306–3309.

77. Christophe BO, Obert B, Meyer D, Girma J. The binding domain of von Willebrand factor to sulfatides is distinct from those interacting with glycoprotein Ib, heparin, and collagen and resides between amino acid residues leu 512 and lys 673. *Blood*. 1991;78(9):2310–2317.

78. Nakayama T, Matsushita T, Yamamoto K, et al. Identification of amino acid residues responsible for von Willebrand factor binding to sulfatide by charged-to-alanine-scanning mutagenesis. *International Journal of Hematology*. 2008;87(4):363–370.

79. Matsushita T, Meyer D, Sadler JE. Localization of von Willebrand factor-binding sites for platelet glycoprotein Ib and botrocetin by charged-to-alanine scanning mutagenesis. *The Journal of Biological Chemistry*. 2000;275(15):11044–11049.

80. Fukuda K, Doggett T, Laurenzi IJ, Liddington RC, Diacovo TG. The snake venom protein botrocetin acts as a biological brace to promote dysfunctional platelet aggregation. *Nature Structural & Molecular Biology*. 2005;12(2):152–159.

81. Han J-H, Batey S, Nickson A a, Teichmann S a, Clarke J. The folding and evolution of multidomain proteins. *Nature Reviews Molecular Cell Biology*. 2007;8(4):319–330.

82. Patthy L. Exon shuffling and other ways of module exchange. *Matrix biology journal of the International Society for Matrix Biology*. 1996;15(5):301–310.

83. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics*. 2010;11(5):345–355.

84. Kolkman JA, Stemmer WP. Directed evolution of proteins by exon shuffling. *Nature biotechnology*. 2001;19(5):423–8.

85. Apic G, Gough J, Teichmann SA. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *Journal of Molecular Biology*. 2001;310(2):311–325.

86. Lang T, Hansson GC, Samuelsson T. Gel-forming mucins appeared early in metazoan evolution. *Proceedings of the National Academy of Sciences*. 2007;104(41):16209–16214.

87. Voorberg J, Fontijn R, Van Mourik JA, Pannekoek H. Domains involved in multimer assembly of von Willebrand factor (VWF): multimerization is independent of dimerization. *The EMBO journal*. 1990;9(3):797–803.

88. Goto A, Kumagai T, Kumagai C, et al. A drosophila haemocyte-specific protein, hemolectin, similar to human von Willebrand factor. *The Biochemical journal*. 2001;359(Pt 1):99–108.

89. Perez-vilar J, Hill RL. Identification of the half-cystine residues in porcine submaxillary mucin critical for multimerization through the D-domains. *Journal of Biological Chemistry*. 1998;273(51):34527–34534.

90. Desseyn J. Genomic organization of the human mucin gene MUC5b. cDNA and genomic sequences upstream of the large central exon. *Journal of Biological Chemistry*. 1998;273(46):30157–30164.

91. Wagner DD, Mayadas T, Marder VJ. Initial glycosylation and acidic pH in the golgi apparatus are required for multimerization of von Willebrand factor. *The Journal of Cell Biology*. 1986;102(4):1320–1324.

92. Purvis AR, Sadler JE. A covalent oxidoreductase intermediate in propeptide-dependent von Willebrand factor multimerization. *The Journal of Biological Chemistry*. 2004;279(48):49982–49988.

93. Mayadas TN, Wagner DD. Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. *Proceedings of the National Academy of Sciences of the United States of America*. 1992;89(8):3531–3535.

94. Vogel C, Berzuini C, Bashton M, Gough J, Teichmann SA. Supra-domains: evolutionary units larger than single protein domains. *Journal of Molecular Biology*. 2004;336(3):809–23.

95. Davidson CJ, Tuddenham EG, McVey JH. 450 million years of hemostasis. *Journal of thrombosis and haemostasis*. 2003;1(7):1487–1494.

96. Holbourn KP, Acharya KR, Perbal B. The CCN family of proteins: structure–function relationships. *Trends in Biochemical Sciences* . 2008;33(10):461–473.

97. Bragdon B, Moseychuk O, Saldanha S, et al. Bone morphogenetic proteins: a critical review. *Cellular Signalling*. 2011;23(4):609–620.

98. Zhang JL, Huang Y, Qiu L-Y, Nickel J, Sebald W. Von willebrand factor type c domain-containing proteins regulate bone morphogenetic protein signaling through different recognition mechanisms. *The Journal of Biological Chemistry*. 2007;282(27):20002–20014.

99. Garcia Abreu J, Coffinier C, Larraín J, Oelgeschläger M, De Robertis EM. Chordin-like cr domains and the regulation of evolutionarily conserved extracellular signaling systems. *Gene*. 2002;287(1-2):39–47.

100. Armishaw CJ. Synthetic α-conotoxin mutants as probes for studying nicotinic acetylcholine receptors and in the development of novel drug leads. *Toxins*. 2010;2(6):1471–1499.

101. Zhang JL, Qiu L-Y, Kotzsch A, et al. Crystal structure analysis reveals how the chordin family member crossveinless 2 blocks bmp-2 receptor binding. *Developmental Cell*. 2008;14(5):739–750.

102. Beacham DA, Wise RJ, Turci SM, Handin RI. Selective inactivation of the arg-gly-asp-ser (rgds) binding site in von Willebrand factor by site-directed mutagenesis. *The Journal of biological chemistry*. 1992;267(5):3409–15.

103. Edman J, Ellis L, Blacher R, Roth R, Rutter W. Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin. *Nature*. 1985;317:267–270.

104. Duggan B, Dyson H, Wright P. Inherent flexibility in a potent inhibitor of blood coagulation, recombinant nematode anticoagulant protein C2. *European journal of biochemistry / FEBS*. 1999;265(2):539–548.

105. Rios-Steiner JL, Murakami MT, Tulinsky A, Arni RK. Active and exo-site inhibition of human factor Xa: structure of des-gla factor Xa inhibited by nap5, a potent nematode anticoagulant protein from ancylostoma caninum. *Journal of Molecular Biology*. 2007;371(3):774–786.

106. Mousa SA. Anticoagulants, antiplatelets, and thrombolytics. Totowa, NJ: Humana Press; 2010.

107. Rawlings ND, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*. 2012;40(Database issue):D343–50.

108. Rawlings ND, Tolle DP, Barrett AJ. Evolutionary families of peptidase inhibitors. *The Biochemical journal*. 2004;378(Pt 3):705–716.

109. Goodman R, Martzen M, Peanasky R. Trypsin inhibitors from *Ascaris*: the reactive P1 site of the inhibitors (a correction) and location of the inhibitors and host trypsin in cross-sections of ascaris. *Acta Biochim Pol*. 1983;30(2):233–244.

110. Bernard V, Peanasky R. The serine protease inhibitor family from *Ascaris suum*: chemical determination of the five disulfide bridges. *Archives of biochemistry and biophysics*. 1993;303(2):367–376.

111. Laskowski M, Kato I. Protein inhibitors of proteinases. *Annual review of biochemistry*. 1980;49:593–626.

112. Laskowski M, Qasim M. What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochimica et biophysica acta*. 2000;1477(1-2):324–337.

113. Krowarsch D, Cierpicki T, Jelen F, Otlewski J. Canonical protein inhibitors of serine proteases. *Cellular and molecular life sciences CMLS*. 2003;60(11):2427–2444.

114. Rice P, Longden I, Bleasby A. EMBOSS: the european molecular biology open software suite. *Trends in genetics : TIG*. 2000;16(6):276–7.

115. Chung S, Subbiah S. A structural explanation for the twilight zone of protein sequence homology. *Structure*. 1996;4(10):1123–1127.

116. Hunt L, Barker W. von Willebrand factor shares a distinctive cysteine-rich domain with thrombospondin and procollagen. *Biochemical and Biophysical Research Communications*. 1987;144(2):876–882.

117. Bork P. The modular architecture of a new family of growth regulators related to connective tissue growth factor. *FEBS letters*. 1993;327(2):125–130.

118. O'Leary JM, Hamilton JM, Deane CM, et al. Solution structure and dynamics of a prototypical chordin-like cysteine-rich repeat (von Willebrand factor type C module) from collagen IIA. *The Journal of Biological Chemistry*. 2004;279(51):53857–53866.

119. Parrini C, Taddei N, Ramazzotti M, et al. Glycine residues appear to be evolutionarily conserved for their ability to inhibit aggregation. *Structure*. 2005;13(8):1143–1151.

120. Steward A, Adhya S, Clarke J. Sequence conservation in Ig-like domains: the role of highly conserved proline residues in the fibronectin type III superfamily. *Journal of molecular biology*. 2002;318(4):935–940.

121. Richardson JS, Richardson DC. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proceedings of the National Academy of Sciences*. 2002;99(5):2754–2759.

122. Wright C, Teichmann S, Clarke J, Dobson C. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature*. 2005;438(7069):878–81.

123. Reshef D, Itzhaki Z, Schueler-Furman O. Increased sequence conservation of domain repeats in prokaryotic proteins. *Trends in Genetics*. 2010;26(9):383–7.

124. Sadowski M, Jones D. The sequence–structure relationship and protein function prediction. *Current opinion in structural biology*. 2009;19(3):357–362.

125. Birzele F, Csaba G, Zimmer R. Alternative splicing and protein structure evolution. *Nucleic acids research*. 2008;36(2):550–558.

126. Romero P, Zaidi S, Fang Y, et al. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proceedings of the National Academy of Sciences*. 2006;103(22):8390–8395.

127. Light S, Elofsson A. The impact of splicing on protein domain architecture. *Current opinion in structural biology*. 2013;In press.

128. Moake J, Turner N, Stathopoulos N, Nolasco L, Hellums J. Involvement of large plasma von willebrand factor (VWF) multimers and unusually large vwf forms derived from endothelial cells in shear stress-induced platelet aggregation. *Journal of Clinical Investigation*. 1986;78(6):1456–1461.

129. Counts RB, Paskell SL, Elgee SK. Disulfide bonds and the quaternary structure of factor VIII/von Willebrand factor. *Journal of Clinical Investigation*. 1978;62(3):702–709.

130. Ruggeri Z, Zimmerman T. The complex multimeric composition of factor VIII/von Willebrand factor. *Blood*. 1981;57(6):1140–1143.

131. Wagner D. Cell biology of von willebrand factor. *Annual review of cell biology*. 1990;6:217–246.

132. Girma J-P, Meyer D, Verweij C, Pannekoek H, JJ S. Structure-function relationship of human von Willebrand factor. *Blood*. 1987;70(3):605–611.

133. Samor B, Mazurier C, Goudemand M. Preliminary results on the carbohydrate moiety of factor viii/von willebrand factor (FVIII/VWF). *Thrombosis Research*. 1982;25(1):81–89.

134. McKinnon T, Chion A, Millington A, Lane D, Laffan M. N-linked glycosylation of VWF modulates its interaction with ADAMTS13. *Blood*. 2008;106(8):2744–2749.

135. Millar C, Brown S. Oligosaccharide structures of von Willebrand factor and their potential role in von willebrand disease. *Blood reviews*. 2006;20(2):83–92.

136. Singh I, Shankaran H, Beauharnois ME, et al. Solution structure of human von Willebrand factor studied using small angle neutron scattering. *Journal of Biological Chemistry*. 2006;281(50):38266–75.

137. Slayters H, Loscalzo J, Bockenstedt P, Handin RI. Native conformation of human von Willebrand protein. analysis by electron microscopy and quasi-elastic light scattering. *Journal of Biological Chemistry*. 1985;260(14):8559–8563.

138. Siedlecki CA, Lestini BJ, Kottke-Marchant KK, et al. Shear-dependent changes in the three-dimensional structure of human von Willebrand factor. *Blood*. 1996;88(8):2939–2950.

139. Fowler WE, Fretto LJ, Hamilton KK, Erickson HP, McKee P a. Substructure of human von Willebrand factor. *Journal of Clinical Investigation*. 1985;76(4):1491–1500.

140. Zhou YF, Eng ET, Nishida N, et al. A pH-regulated dimeric bouquet in the structure of von Willebrand factor. *the The European Molecular Biology Organization Journal*. 2011;30(19):4098–4111.

141. Denis C. Molecular and cellular biology of von Willebrand factor. *International journal of hematology*. 2002;75(1):3–8.

142. Canis K, McKinnon T a J, Nowak A, et al. Mapping the N-glycome of human von Willebrand factor. *The Biochemical journal*. 2012;447(2):217–28.

143. Matsui T, Titani K, Mizuochi T. Structures of the asparagine-linked oligosaccharide chains of human von Willebrand factor. *Journal of Biological Chemistry*. 1992;267(13):8723–8731.

144. Matsui T, Fujimura Y, Nishida S, Titani K. Human plasma alpha 2-macroglobulin and von Willebrand factor possess covalently linked ABO(H) blood group antigens in subjects with corresponding ABO phenotype. *Blood*. 1993;82(2):663–668.

145. McKinnon TAJ, Goode EC, Birdsey GM, et al. Specific N-linked glycosylation sites modulate synthesis and secretion of von Willebrand factor. *Blood*. 2010;116(4):640–648.

146. Jenkins PV, O'Donnell JS. ABO blood group determines plasma von Willebrand factor levels: a biologic function after all? *Transfusion*. 2006;46(10):1836–44.

147. Lenting PJ, VAN Schooten CJM, Denis CV. Clearance mechanisms of von Willebrand factor and factor viii. *Journal of thrombosis and haemostasis : JTH*. 2007;5(7):1353–60.

148. O'Donnell JS, McKinnon TAJ, Crawley JTB, Lane DA, Laffan MA. Bombay phenotype is associated with reduced plasma-VWF levels and an increased susceptibility to ADAMTS13 proteolysis. *Blood*. 2005;106(6):1988–91.

149. Bowen D. An influence of ABO blood group on the rate of proteolysis of von Willebrand factor by ADAMTS13. *Journal of Thrombosis and Haemostasis*. 2003;1(1):33–40.

150. McGrath RT, McKinnon TAJ, Byrne B, et al. Expression of terminal alpha2-6-linked sialic acid on von Willebrand factor specifically enhances proteolysis by ADAMTS13. *Blood*. 2010;115(13):2666–2673.

151. James PD, Notley C, Hegadorn C, et al. The mutational spectrum of type 1 von Willebrand disease: results from a Canadian cohort study. *Blood*. 2007;109(1):145–154.

152. Katsumi A, Tuley EA, Bodo I, Sadler JE. Localization of disulfide bonds in the cystine knot domain of human von Willebrand factor. *Journal of Biological Chemistry*. 2000;275(33):25585–25594.

153. Meitinger T, Meindl A, Bork P, et al. Molecular modelling of the norrie disease protein predicts a cystine knot growth factor tertiary structure. *Nature genetics*. 1993;5(4):376–380.

154. Tjernberg P, Vos HL, Spaargaren-van Riel CC, et al. Differential effects of the loss of intrachain- versus interchain-disulfide bonds in the cystine-knot domain of von willebrand factor on the clinical phenotype of von Willebrand disease. *Thrombosis and haemostasis*. 2006;96(6):717–724.

155. Schneppenheim R, Brassard J, Krey S, et al. Defective dimerization of von Willebrand factor subunits due to a Cys->Arg mutation in type IID von Willebrand disease. *Proceedings of the National Academy of Sciences*. 1996;93(8):3581–3586.

156. Sadler JE. von Willebrand factor assembly and secretion. *Journal of thrombosis and haemostasis JTH*. 2009;7 Suppl 1(s1 State of the Art 2009):24–27.

157. Carew J, Browning P, Lynch D. Sulfation of von Willebrand factor. *Blood*. 1990;76(3):2530–2539.

158. Canis K, McKinnon TAJ, Nowak A, et al. The plasma von willebrand factor o-glycome comprises a surprising variety of structures including abh antigens and disialosyl motifs. *Journal of thrombosis and haemostasis JTH*. 2010;8(1):137–145.

159. Thomas G. Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nature Reviews Molecular Cell Biology*. 2002;3(10):753–66.

160. 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):229–236.

161. Verweij CL, Hart M, Pannekoek H. Proteolytic cleavage of the precursor of von Willebrand factor is not essential for multimer formation. *The Journal of Biological Chemistry*. 1988;263(17):7921–7924.

162. Hilbert L, Nurden P, Caron C, 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. *Thrombosis and haemostasis*. 2006;96(3):290–294.

163. Sevier CS, Kaiser CA. Formation and transfer of disulphide bonds in living cells. *Nature Reviews Molecular Cell Biology*. 2002;3(11):836–847.

164. Mayadas TN, Wagner DD. In vitro multimerization of von Willebrand factor is triggered by low pH: importance of the propolypeptide and free sulfhydryls. *The Journal of Biological Chemistry*. 1989;264(23):13497–13503.

165. Seksek O, Biwersi J, Verkman A. Direct measurement of trans-golgi pH in living cells and regulation by second messengers. *Journal of Biological Chemistry*. 1995;270(10):4967–4970.

166. Huang R-H, Wang Y, Roth R, et al. Assembly of Weibel-Palade body-like tubules from N-terminal domains of von Willebrand factor. *Proceedings of the National Academy of Sciences*. 2008;105(2):482–487.

167. Ruggeri Z. Structure and function of von Willebrand factor. *Thrombosis and haemostasis*. 1999;82(2):576–584.

168. Van Schooten CJ, Shahbazi S, Groot E, et al. Macrophages contribute to the cellular uptake of von Willebrand factor and factor VIII in vivo. *Blood*. 2008;112(5):1704–12.

169. Berriman JA, Li S, Hewlett LJ, et al. Structural organization of weibel-palade bodies revealed by cryo-em of vitrified endothelial cells. *Proceedings of the National Academy of Sciences*. 2009;106(41):17407–17412.

170. Valentijn KM, Eikenboom J. Weibel-Palade bodies: a window to von Willebrand disease. *Journal of thrombosis and haemostasis : JTH*. 2013;11(4):581–92.

171. Erent M, Meli A, Moisoi N, et al. Rate, extent and concentration dependence of histamine-evoked Weibel-Palade body exocytosis determined from individual fusion events in human endothelial cells. *The Journal of Physiology*. 2007;583(Pt 1):195–212.

172. Springer T. Biology and physics of von Willebrand factor concatamers. *Journal of Thrombosis and Haemostasis*. 2011;9(Suppl 1):130–143.

173. Zenner HL, Collinson LM, Michaux G, Cutler DF. High-pressure freezing provides insights into Weibel-Palade body biogenesis. *Journal of cell science*. 2007;120(12):2117–25.

174. Luo G-P, Ni B, Yang X, Wu Y-Z. Von Willebrand factor: more than a regulator of hemostasis and thrombosis. *Acta haematologica*. 2012;128(3):158–69.

175. Kaufmann JE, Vischer UM. Cellular mechanisms of the hemostatic effects of desmopressin (DDAVP). *Journal of thrombosis and haemostasis JTH*. 2003;1(4):682–689.

176. Huang J, Roth R, Heuser JE, Sadler JE. Integrin αvβ3 on human endothelial cells binds von Willebrand factor strings under fluid shear stress. *Blood*. 2009;113(7):1589–1597.

177. Dong J, Moake JL, Nolasco L, et al. ADAMTS-13 rapidly cleaves newly secreted ultralarge von Willebrand factor multimers on the endothelial surface under flowing conditions. *Blood*. 2002;100(12):4033–4039.

178. Ganderton T, Wong JWH, 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.

179. Yeh HC, Zhou Z, Choi H, et al. Disulfide bond reduction of von Willebrand factor by ADAMTS-13. *Journal of thrombosis and haemostasis JTH*. 2010;8(12):2778–2788.

180. Zhang X, Halvorsen K, Zhang C, Wong W, Springer T. Mechanoenzymatic cleavage of the ultralarge vascular protein von Willebrand factor. *Science*. 2009;324(5932):1330–1334.

181. Arya M, Anvari B, Romo GM, et al. Ultralarge multimers of von Willebrand factor form spontaneous high-strength bonds with the platelet glycoprotein Ib-IX complex: studies using optical tweezers. *Blood*. 2002;99(11):3971–3977.

182. De Ceunynck K, De Meyer SF, Vanhoorelbeke K. Unwinding the von willebrand factor strings puzzle. *Blood*. 2013;121(2):270–7.

183. Moake J, Rudy C, Troll J, et al. Unusually large plasma factor VIII: von willebrand factor multimers in chronic relapsing thrombotic thrombocytopenic purpura. *The New England Journal of Medicine*. 1982;307(23):1432–1435.

184. Tsai H-M. Pathophysiology of thrombotic thrombocytopenic purpura. *International Journal of Hematology*. 2010;91(1):1–19.

185. Sadler JE. Von Willebrand factor, ADAMTS13, and thrombotic thrombocytopenic purpura. *Blood*. 2008;112(1):11–18.

186. Levy GG, Nichols WC, Lian EC, et al. Mutations in a member of the adamts gene family cause thrombotic thrombocytopenic purpura. *Nature*. 2001;413(6855):488–494.

187. Batlle J, Lopez-Fernandez M, Lopez-Borrasca A, et al. Proteolytic degradation of von willebrand factor after DDAVP administration in normal individuals. *Blood*. 1987;70(1):173–176.

188. Tsai HM. Physiologic cleavage of von willebrand factor by a plasma protease is dependent on its conformation and requires calcium ion. *Blood*. 1996;87(10):4235–4244.

189. Furlan M, Robles R, Lamie 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):4223–4234.

190. Zhang Q, Zhou YF, Zhang CZ, et al. Structural specializations of A2, a force-sensing domain in the ultralarge vascular protein von Willebrand factor. *Proceedings of the National Academy of Sciences*. 2009;106(23):9226–9231.

191. Wang A, Liu F, Dong N, et al. Thrombospondin-1 and ADAMTS13 competitively bind to Vwf A2 and A3 domains in vitro. *Thrombosis Research*. 2010;126(4):e260–e265.

192. Sadler JE, Budde U, Eikenboom JCJ, et al. Update on the pathophysiology and classification of von Willebrand disease: a report of the subcommittee on von willebrand factor. *Journal of thrombosis and haemostasis JTH*. 2006;4(10):2103–2114.

193. Xu AJ, Springer TA. Mechanisms by which von willebrand disease mutations destabilize the A2 domain. *The Journal of biological chemistry*. 2013;288(9):6317–24.

194. 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. *The Journal of Biological Chemistry*. 1998;273(17):10396–401.

195. 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):1147–1156.

196. 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. *The Journal of Biological Chemistry*. 1997;272(40):25162–25167.

197. Brondijk THC, Bihan D, Farndale RW, Huizinga EG. Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3

domain complex. *Proceedings of the National Academy of Sciences*. 2012;109(14):5253–5258.

198. Hathcock JJ. Flow effects on coagulation and thrombosis. *Arteriosclerosis, thrombosis, and vascular biology*. 2006;26(8):1729–37.

199. Schneider SW, Nuschele S, Wixforth A, et al. Shear-induced unfolding triggers adhesion of von Willebrand factor fibers. *Proceedings of the National Academy of Sciences*. 2007;104(19):7899–7903.

200. Alexander-Katz A, Schneider MF, Schneider SW, Wixforth A, Netz RR. Shear-flow-induced unfolding of polymeric globules. *Physical Review Letters*. 2006;97(13):138101.

201. Steppich DM, Angerer JI, Sritharan K, et al. Relaxation of ultralarge VWF bundles in a microfluidic-AFM hybrid reactor. *Biochemical and Biophysical Research Communications*. 2008;369(2):507–512.

202. Back L, Radbill J, Crawford D. Analysis of pulsatile, viscous blood flow through diseased coronary arteries of man. *Journal of biomechanics*. 1977;10(5):339–353.

203. De Meyer SF, Stoll G, Wagner DD, Kleinschnitz C. Von Willebrand factor: an emerging target in stroke therapy. *Stroke*. 2012;43(2):599–606.

204. Lenting PJ, Casari C, Christophe OD, Denis C V. Von willebrand factor: the old, the new and the unknown. *Journal of thrombosis and haemostasis : JTH*. 2012;10(12):2428–2437.

205. Denis C V, Lenting PJ. von Willebrand factor: at the crossroads of bleeding and thrombosis. *International Journal of Hematology*. 2012;95(4):353–61.

206. Vischer U. von Willebrand factor, endothelial dysfunction, and cardiovascular disease. *Journal of Thrombosis and Haemostasis*. 2006;4(6):1186–1193.

207. Baruch D, Denis C, Marteaux C, et al. Role of von Willebrand factor associated to extracellular matrices in platelet adhesion. *Blood*. 1991;77(3):519–527.

208. Rand JH, Wu XX, Potter BJ, Uson RR, Gordon RE. Co-localization of von Willebrand factor and type VI collagen in human vascular subendothelium. *The American journal of pathology*. 1993;142(3):843–850.

209. Lankhof H, Van Hoeij M, Schiphorst M, et al. A3 domain is essential for interaction of von Willebrand factor with collagen type III. *Thrombosis and haemostasis*. 1996;75(6):950.

210. Cruz M, Yuan H, Lee J, Wise R, Handin R. Interaction of the von Willebrand factor (VWF) with collagen. localization of the primary collagen-binding site by analysis

of recombinant VWF A domain polypeptides. *Journal of Biological Chemistry*. 1995;270(18):10822–10827.

211. Ribba A, Loisel I, Lavergne JM, et al. Ser968Thr mutation within the A3 domain of von Willebrand factor (VWF) in two related patients leads to a defective binding of VWF to collagen. *Thrombosis and haemostasis*. 2001;86(3):848–854.

212. Riddell AF, Gomez K, Millar CM, et al. Characterization of W1745C and S1783A: 2 novel mutations causing defective collagen binding in the A3 domain of von Willebrand factor. *Blood*. 2009;114(16):3489–3496.

213. Flood VH, Lederman C a, Wren JS, et al. Absent collagen binding in a VWF A3 domain mutant: utility of the VWF:CB in diagnosis of VWD. *Journal of thrombosis and haemostasis : JTH*. 2010;8(6):1431–3.

214. Bonnefoy A, Romijn RA, Vandervoort PAH, et al. Von willebrand factor A1 domain can adequately substitute for A3 domain in recruitment of flowing platelets to collagen. *Journal of thrombosis and haemostasis JTH*. 2006;4(10):2151–2161.

215. Guidetti GF, Bartolini B, Bernardi B, et al. Binding of von Willebrand factor to the small proteoglycan decorin. *FEBS Letters*. 2004;574(1-3):95–100.

216. Lisman T, Raynal N, Groeneveld D, 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):3753–3756.

217. Lasser G, Guchhait P, Ellsworth JL, et al. C1qtnf-related protein-1 (ctrp-1): a vascular wall protein that inhibits collagen-induced platelet aggregation by blocking VWF binding to collagen. *Blood*. 2006;107(2):423–430.

218. Wu D, Vanhoorelbeke K, Cauwenberghs N, 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):3623–3628.

219. Savage B, Saldívar E, Ruggeri ZM. Initiation of platelet adhesion by arrest onto fibrinogen or translocation on von Willebrand factor. *Cell*. 1996;84(2):289–297.

220. Sing CE, Alexander-Katz A. Elongational flow induces the unfolding of von Willebrand factor at physiological flow rates. *Biophysical Journal*. 2010;98(9):L35–L37.

221. 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. *Journal of thrombosis and haemostasis JTH*. 2006;4(2):417–425.

222. 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):2637–2645.

223. Auton M, Sowa KE, Behymer M, Cruz MA. N-terminal flanking region of A1 domain in von Willebrand factor stabilizes structure of A1A2A3 complex and modulates platelet activation under shear stress. *The Journal of biological chemistry*. 2012;287(18):14579–85.

224. Ulrichts H, Udvardy M, Lenting PJ, et al. Shielding of the A1 domain by the D'D3 domains of von Willebrand factor modulates its interaction with platelet glycoprotein IB-IX-V. *Journal of Biological Chemistry*. 2006;281(8):4699–4707.

225. Madabhushi SR, Shang C, Dayananda KM, et al. Von Willebrand factor propeptide (VWFpp) binding to VWF D'D3 domain attenuates platelet activation and adhesion. *Blood*. 2012;119(20):4769–78.

226. Kim J, Zhang C-Z, Zhang X, Springer TA. A mechanically stabilized receptor-ligand flex-bond important in the vasculature. *Nature*. 2010;466(7309):992–995.

227. Yago T, Lou J, Wu T, et al. Platelet glycoprotein Ibα forms catch bonds with human wt VWF but not with type 2B von willebrand disease VWF. *Journal of Clinical Investigation*. 2008;118(9):3195–3207.

228. Doggett T a, Girdhar G, Lawshe A, et al. Alterations in the intrinsic properties of the GPIbalpha-VWF tether bond define the kinetics of the platelet-type von Willebrand disease mutation, Gly233Val. *Blood*. 2003;102(1):152–160.

229. Szántó T, Joutsi-Korhonen L, Deckmyn H, Lassila R. New insights into von Willebrand disease and platelet function. *Seminars in thrombosis and hemostasis*. 2012;38(1):55–63.

230. Rayes J, Hommais A, Legendre P, et al. Effect of von Willebrand disease type 2B and type 2M mutations on the susceptibility of von Willebrand factor to ADAMTS-13. *Journal of thrombosis and haemostasis : JTH*. 2007;5(2):321–8.

231. Berndt MC, Andrews RK. Bernard-Soulier syndrome. *Haematologica*. 2011;96(3):355–9.

232. Berndt MC, Shen Y, Dopheide SM, Gardiner EE, Andrews RK. The vascular biology of the glycoprotein IB-IX-V complex. *Thrombosis and haemostasis*. 2001;86(1):178–188.

233. Reininger A. Function of von Willebrand factor in haemostasis and thrombosis. *Haemophilia*. 2008;14(Suppl 5):11–26.

234. Broos K, Feys HB, De Meyer SF, Vanhoorelbeke K, Deckmyn H. Platelets at work in primary hemostasis. *Blood Reviews*. 2011;25(4):155–167.

235. Franchini M, Favaloro E, Lippi G. Glanzmann thrombasthenia: an update. *Clinica Chimica Acta*. 2010;411(1-2):1–6.

236. Hampshire DJ, Ph D, Goodeve AC. The international society on thrombosis and haematosis von Willebrand disease database : an update. *Seminars In Thrombosis And Hemostasis*. 2011;1(212):470–479.

237. Takahashi Y, Kalafatis M, Girma JP, et al. Localization of a factor VIII binding domain on a 34 kilodalton fragment of the N-terminal portion of von Willebrand factor. *Blood*. 1987;70(5):1679–1682.

238. Pietu G, Ribba A, Meulien P, Meyer D. Localization within the 106 N-terminal amino acids of von Willebrand factor (VWF) of the epitope corresponding to a monoclonal antibody which inhibits VWF binding to. *Biochemical and biophysical Research Communications*. 1989;163(1):618–626.

239. 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):3983–3996.

240. Tuddenham EG, Lane RS, Rotblat F, et al. Response to infusions of polyelectrolyte fractionated human factor VIII concentrate in human haemophilia A and von Willebrand's disease. *British journal of haematology*. 1982;52(2):259–67.

241. Nishino M, Girma JP, Rothschild C, Fressinaud E, Meyer D. New variant of von Willlebrand disease with defective binding to factor VIII. *Blood*. 1989;74(5):1591–1599.

242. Wion KL, Kelly D, Summerfield JA, Tuddenham EGD, Lawn RM. Distribution of factor VIII mRNA and antigen in human liver and other tissues. *Nature*. 1985;317(6039):726–729.

243. Hollestelle MJ, Thinnes T, Crain K, et al. Tissue distribution of factor VIII gene expression in vivo-a closer look. *Thrombosis and haemostasis*. 2001;86(3):855–61.

244. Lamont PA, Ragni MV. Lack of desmopressin (DDAVP) response in men with hemophilia A following liver transplantation. *Journal of thrombosis and haemostasis JTH*. 2005;3(10):2259–2263.

245. Haberichter SL, Shi Q, Montgomery RR. Regulated release of VWF and FVIII and the biologic implications. *Pediatric blood cancer*. 2006;46(5):547–553.

246. Pipe SW. Functional roles of the factor VIII B domain. *Haemophilia*. 2009;15(6):1187–96.

247. Vlot a J, Koppelman SJ, Van Den Berg MH, Bouma BN, Sixma JJ. The affinity and stoichiometry of binding of human factor VIII to von Willebrand factor. *Blood*. 1995;85(11):3150–3157.

248. Lollar P, Hill-Eubanks DC, Parker CG. Association of the factor VIII light chain with von Willebrand factor. *The Journal of biological chemistry*. 1988;263(21):10451–10455.

249. Leyte A, Verbeet M, Brodniewicz-Proba T, Van Mourik J, Mertens K. The interaction between human blood-coagulation factor VIII and von Willebrand factor. characterization of a high-affinity binding site on factor VIII. *Biochemical Journal*. 1989;257(3):679–683.

250. Vlot AJ, Koppelman SJ, Bouma BN, Sixma JJ. Factor VIII and von Willebrand factor. *Thrombosis and haemostasis*. 1998;79(3):456–65.

251. Vlot A, Koppelman SJ, Meijers JC, et al. Kinetics of factor VIII-von Willebrand factor association. *Blood*. 1996;87(5):1809–16.

252. Dimitrov JD, Christophe OD, Kang J, et al. Thermodynamic analysis of the interaction of factor VIII with von Willebrand factor. *Biochemistry*. 2012;51(20):4108–16.

253. Brinkhous KM, Sandberg H, Garris JB, et al. Purified human factor VIII procoagulant protein: comparative hemostatic response after infusions into hemophilic and von willebrand disease dogs. *Proceedings of the National Academy of Sciences*. 1985;82(24):8752–6.

254. Noe DA. A mathematical model of coagulation factor VIII kinetics. *Pathophysiology of Haemostasis and Thrombosis*. 1996;26(6):289–303.

255. Miller CH, Haff E, Platt SJ, et al. Measurement of von Willebrand factor activity: relative effects of ABO blood type and race. *Journal of thrombosis and haemostasis JTH*. 2003;1(10):2191–2197.

256. Koster T, Blann a D, Briët E, Vandenbroucke JP, Rosendaal FR. Role of clotting factor VIII in effect of von willebrand factor on occurrence of deep-vein thrombosis. *Lancet*. 1995;345(8943):152–155.

257. Franchini M, Favaloro EJ, Targher G, Lippi G. ABO blood group, hypercoagulability, and cardiovascular and cancer risk. *Critical reviews in clinical laboratory sciences*. 2012;49(4):137–49.

258. Wood WI, Capon DJ, Simonsen CC, et al. Expression of active human factor VIII from recombinant DNA clones. *Nature*. 1984;312(5992):330–337.

259. Vehar GA, Keyt B, Eaton D, et al. Structure of human factor VIII. *Nature*. 1984;312(5992):337–42.

260. Fay PJ. Factor VIII structure and function. *International journal of hematology*. 2006;83(2):103–8.

261. Mannucci PM, Tuddenham EG. The hemophilias--from royal genes to gene therapy. *The New England journal of medicine*. 2001;344(23):1773–9.

262. Saenko EL, Loster K, Josic D, Sarafanov a G. Effect of von Willebrand factor and its proteolytic fragments on kinetics of interaction between the light and heavy chains of human factor VIII. *Thrombosis research*. 1999;96(5):343–54.

263. Hamer RJ, Koedam JA, Beeser-Visser NH, et al. Factor VIII binds to von Willebrand factor via its Mr-80,000 light chain. *European journal of biochemistry / FEBS*. 1987;166(1):37–43.

264. 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. *The Journal of biological chemistry*. 1997;272(29):18007–18014.

265. Leyte A, Van Schijndel HB, Niehrs C, et al. Sulfation of Tyr1680 of human blood coagulation factor VIII is essential for the interaction of factor VIII with von Willebrand factor. *The Journal of Biological Chemistry*. 1991;266(2):740–746.

266. Fay PJ, Jenkins PV. Mutating factor VIII: lessons from structure to function. *Blood reviews*. 2005;19(1):15–27.

267. Gringeri A, Ofosu FA, Grancha S, et al. Understanding FVIII/VWF complex--report from a symposium of XXIX WFH meeting 2010. *Haemophilia*. 2012;18(3):469–75.

268. Castro-Núñez L, Bloem E, Boon-Spijker MG, et al. Distinct roles of Ser-764 and Lys-773 at the N terminus of von Willebrand factor in complex assembly with coagulation factor VIII. *The Journal of biological chemistry*. 2013;288(1):393–400.

269. Saenko E, Scandella D. A mechanism for inhibition of factor VIII binding to phospholipid by von Willebrand factor. *Journal of Biological Chemistry*. 1995;270(23):13826–13833.

270. Bendetowicz AV, Morris JA, Wise RJ, Gilbert GE, Kaufman RJ. Binding of factor viii to von Willebrand factor is enabled by cleavage of the von Willebrand factor propeptide and enhanced by formation of disulfide-linked multimers. *Blood*. 1998;92(2):529–38.

271. Fay PJ. Activation of factor VIII and mechanisms of cofactor action. *Blood Reviews*. 2004;18(1):1–15.

272. Ngo JCK, Huang M, Roth DA, Furie BC, Furie B. Crystal structure of human factor VIII: implications for the formation of the factor IXa-factor VIIIa complex. *Structure*. 2008;16(4):597–606.

273. Shen BW, Spiegel PC, Chang CH, et al. The tertiary structure and domain organization of coagulation factor VIII. *Blood*. 2008;111(3):1240–1247.

274. Spiegel PC, Murphy P, Stoddard BL. Surface-exposed hemophilic mutations across the factor VIII C2 domain have variable effects on stability and binding activities. *The Journal of biological chemistry*. 2004;279(51):53691–53698.

275. Gilbert GE, Kaufman RJ, Arena AA, Miao H, Pipe SW. Four hydrophobic amino acids of the factor VIII C2 domain are constituents of both the membrane-binding and von Willebrand factor-binding motifs. *The Journal of biological chemistry*. 2002;277(8):6374–81.

276. Dimitrov JD, Roumenina LT, Plantier J-L, et al. A human FVIII inhibitor modulates FVIII surface electrostatics at a VWF-binding site distant from its epitope. *Journal of thrombosis and haemostasis JTH*. 2010;8(7):1524–1531.

277. Gilbert GE, Novakovic VA, Kaufman RJ, Miao H, Pipe SW. Conservative mutations in the C2 domains of factor VIII and factor V alter phospholipid binding and cofactor activity. *Blood*. 2012;120(9):1923–32.

278. Jacquemin M, Lavend'homme R, Benhida A, et al. A novel cause of mild/moderate hemophilia A: mutations scattered in the factor VIII C1 domain reduce factor VIII binding to von Willebrand factor. *Blood*. 2000;96(3):958–65.

279. Jacquemin M, Benhida A, Peerlinck K, et al. A human antibody directed to the factor VIII C1 domain inhibits factor VIII cofactor activity and binding to von Willebrand factor a human antibody directed to the factor VIII C1 domain inhibits factor VIII cofactor activity and binding to von Willebrand. *Blood*. 2000;95(1):156–163.

280. Lü J, Pipe SW, Miao H, Jacquemin M, Gilbert GE. A membrane-interactive surface on the factor VIII C1 domain cooperates with the C2 domain for cofactor function. *Blood*. 2011;117(11):3181–9.

281. Liu ML, Shen BW, Nakaya S, et al. Hemophilic factor VIII C1- and C2-domain missense mutations and their modeling to the 1.5-angstrom human C2-domain crystal structure. *Blood*. 2000;96(3):979–987.

282. Venceslá A, Corral-Rodríguez MA, Baena M, et al. Identification of 31 novel mutations in the F8 gene in spanish hemophilia a patients: structural analysis of 20 missense mutations suggests new intermolecular binding sites. *Blood*. 2008;111(7):3468–78.

283. Cooper HA, Griggs TR, Wagner RH. Factor VIII recombination after dissociation by $CaCl_2$. *Proceedings of the National Academy of Sciences*. 1973;70(8):2326–2329.

284. Weiss HJ, Hoyer LW. Von Willebrand factor : dissociation from antihemophilic factor procoagulant activity. *Science*. 1973;182(4117):1149–1151.

285. Schreiber G, Haran G, Zhou H-X. Fundamental aspects of protein-protein association kinetics. *Chemical reviews*. 2009;109(3):839–60.

286. Favaloro EJ, Mohammed S, Koutts J. Identification and prevalence of von Willebrand disease type 2N (Normandy) in Australia. *Blood coagulation & fibrinolysis*. 2009;20(8):706–14.

287. Favaloro EJ. Interaction of factor VIII and von Willebrand factor and the identification of type 2N von willebrand disease. *Thrombosis research*. 2011;127(1):2–3.

288. Casonato A, Daidone V, Barbon G, et al. A common ancestor more than 10,000 years old for patients with R854Q-related type 2N von willebrand's disease in italy. *Haematologica*. 2013;98(1):147–52.

289. Koppelman SJ, Van Hoeij M, Vink T, et al. Requirements of von Willebrand factor to protect factor viii from inactivation by activated protein C. *Blood*. 1996;87(6):2292–300.

290. Wise RJ, Dorner AJ, Krane M, Pittman DD, Kaufman RJ. The role of von Willebrand factor multimers and propeptide cleavage in binding and stabilization of factor VIII. *The Journal of biological chemistry*. 1991;266(32):21948–55.

291. Fay P. Reconstitution of human factor VIII from isolated subunits. *Archives of biochemistry and biophysics*. 1988;262(2):525–531.

292. Koedam JA, Meijers JC, Sixma JJ, Bouma BN. Inactivation of human factor VIII by activated protein C. cofactor activity of protein S and protective effect of von Willebrand factor. *The Journal of clinical investigation*. 1988;82(4):1236–43.

293. 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. *European Journal of Biochemistry*. 1990;189(2):229–234.

294. Fay PJ, Coumans J V, Walker FJ. Von willebrand factor mediates protection of factor VIII from activated protein C-catalyzed inactivation. *The Journal of biological chemistry*. 1991;266(4):2172–7.

295. Lenting PJ, Donath MJ, Van Mourik JA, Mertens K. Identification of a binding site for blood coagulation factor IXa on the light chain of human factor VIII. *The Journal of biological chemistry*. 1994;269(10):7150–5.

296. Lenting PJ, Christophe OD, Guéguen P. The disappearing act of factor VIII. *Haemophilia*. 2010;16(102):6–15.

297. Lenting PJ, Van de Loo JW, Donath MJ, Van Mourik JA, Mertens K. The sequence Glu1811-Lys1818 of human blood coagulation factor VIII comprises a binding site for activated factor IX. *The Journal of biological chemistry*. 1996;271(4):1935–40.

298. Nesheim M, Pittman DD, Giles AR, et al. The effect of plasma von Willebrand factor on the binding of human factor VIII to thrombin-activated human platelets. *The Journal of biological chemistry*. 1991;266(27):17815–20.

299. Nogami K, Shima M, Nishiya K, et al. A novel mechanism of factor VIII protection by von Willebrand factor from activated protein C-catalyzed inactivation. *Blood*. 2002;99(11):3993–8.

300. Gilbert E, Furie B, Furie B. Binding of human factor VIII to phospholipid vesicles. *Journal of Biological Chemistry*. 1990;265(2):815–822.

301. Lenting PJ, Neels JG, Van den Berg BM, et al. The light chain of factor VIII comprises a binding site for low density lipoprotein receptor-related protein. *The Journal of biological chemistry*. 1999;274(34):23734–23739.

302. Saenko EL, Yakhyaev AV, Mikhailenko I, Strickland DK, Sarafanov AG. Role of the low density lipoprotein-related protein receptor in mediation of factor VIII catabolism. *The Journal of biological chemistry*. 1999;274(53):37685–37692.

303. Castro-Núñez L, Dienava-Verdoold I, Herczenik E, Mertens K, Meijer AB. Shear stress is required for the endocytic uptake of the factor VIII-von Willebrand factor complex by macrophages. *Journal of thrombosis and haemostasis : JTH*. 2012;10(9):1929–37.

304. Rastegarlari G, Pegon JN, Casari C, et al. Macrophage LRP1 contributes to the clearance of von Willebrand factor. *Blood*. 2012;119(9):2126–34.

305. Ragni M V. VWF: factor VIII protector and friend. *Journal of thrombosis and haemostasis : JTH*. 2012;2324–2327.

306. Scandella D, Gilbert GE, Shima M, et al. Some factor VIII inhibitor antibodies recognize a common epitope corresponding to C2 domain amino acids 2248 through 2312, which overlap a phospholipid-binding site. *Blood*. 1995;86(5):1811–9.

307. Powell JS, Josephson NC, Quon D, et al. Safety and prolonged activity of recombinant factor VIII Fc fusion protein in hemophilia a patients. *Blood*. 2012;119(13):3031–7.

308. Pipe SW. The hope and reality of long-acting hemophilia products. *American journal of hematology*. 2012;87 Suppl 1(February):S33–9.

309. Mittermaier A, Kay LE. New tools provide new insights in NMR studies of protein dynamics. *Science*. 2006;312(5771):224–8.

310. Schneppenheim R, Budde U, Krey S, et al. Results of a screening for von Willebrand disease type 2N in patients with suspected haemophilia a or von Willebrand disease type 1. *Thrombosis and haemostasis*. 1996;76(4):598–602.

311. Mazurier C, Goudemand J, Hilbert L, et al. Type 2N von Willebrand disease: clinical manifestations, pathophysiology, laboratory diagnosis and molecular biology. *Best practice research Clinical haematology*. 2001;14(2):337–347.

312. James PD, Lillicrap D. Von Willebrand disease: clinical and laboratory lessons learned from the large von Willebrand disease studies. *American Journal of Hematology*. 2012;87 Suppl 1(February):S4–11.

313. Piétu G, Ribba A, Chérel G, et al. Epitope mapping of inhibitory monoclonal antibodies to human von Willebrand factor by using recombinant cDNA libraries. *Thrombosis and haemostasis*. 1994;71(6):788–792.

314. Jorieux S, Gaucher C, Pietu G, Cherel G, Meyer D. Fine epitope mapping of monoclonal antibodies to the NH2-terminal part of von Willebrand factor (VWF) by using recombinant and synthetic peptides: interest for the localization of the factor VIII binding domain. *British Journal of Haematology*. 1994;87:113–118.

315. Hilbert L, Jorieux S, Fontenay-Roupie M, et al. Expression of two type 2N von Willebrand disease mutations identified in exon 18 of von Willebrand factor gene. *British Journal of Haematology*. 2004;127(2):184–189.

316. Huang K, Strynadka NC, Bernard VD, Peanasky RJ, James MN. The molecular structure of the complex of Ascaris chymotrypsin/elastase inhibitor with porcine elastase. *Structure*. 1994;2(7):679–89.

317. Cierpicki T, Bania J, Otlewski J. NMR solution structure of apis mellifera chymotrypsin/cathepsin g inhibitor-1 (AMCI-1): structural similarity with Ascaris protease inhibitors. *Protein Science*. 2000;9(5):976–984.

318. Grasberger BL, Clore GM, Gronenborn a M. 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):669–678.

319. Rosengren KJ, Daly NL, Scanlon MJ, Craik DJ. Solution structure of BSTI: a new trypsin inhibitor from skin secretions of bombina bombina. *Biochemistry*. 2001;40(15):4601–9.

320. Holm L, Rosenström P. Dali server: conservation mapping in 3D. *Nucleic acids research*. 2010;38(Suppl 2):W545–W549.

321. Petersen MT, Jonson PH, Petersen SB. Amino acid neighbours and detailed conformational analysis of cysteines in proteins. *Protein engineering*. 1999;12(7):535–48.

322. Thornton JM. Disulphide bridges in globular proteins. *Journal of Molecular Biology*. 1981;151(2):261–287.

323. Williams M, Phan I, Harvey T, et al. Solution structure of a pair of fibronectin type 1 modules with fibrin binding activity. *Journal of Molecular Biology*. 1994;235(4):1302–1311.

324. Feige MJ, Hendershot LM. Disulfide bonds in er protein folding and homeostasis. *Current opinion in cell biology*. 2011;23(2):167–75.

325. Fass D. Disulfide bonding in protein biophysics. *Annual review of biophysics*. 2012;41:63–79.

326. Wong JWH, Ho SYW, Hogg PJ. Disulfide bond acquisition through eukaryotic protein evolution. *Molecular biology and evolution*. 2011;28(1):327–34.

327. Rubinstein R, Fiser A. Predicting disulfide bond connectivity in proteins by correlated mutations analysis. *Bioinformatics*. 2008;24(4):498–504.

328. Sheldon TJ, Miguel-Aliaga I, Gould a P, Taylor WR, Conklin D. A novel family of single VWC-domain proteins in invertebrates. *FEBS Letters*. 2007;581(27):5268–5274.

329. Cheek S, Krishna SS, Grishin N V. Structural classification of small, disulfide-rich protein domains. *Journal of molecular biology*. 2006;359(1):215–37.

330. Harrison P, Sternberg M. Analysis and classification of disulphide connectivity in proteins: the entropic effect of cross-linkage. *Journal of molecular biology*. 1994;244:448–463.

331. Chuang C, Chen C, Yang J, Lyu P, Hwang J. Relationship between protein structures and disulfide- bonding patterns. 2003;5(April):1–5.

332. Gupta A, Vlijmen H Van, Singh J. A classification of disulfide patterns and its relationship to protein structure and function. *Protein science*. 2004;13(8):2045–2058.

333. Panchenko A, Wolf Y, Panchenko L, Madej T. Evolutionary plasticity of protein families: coupling between sequence and structure variation. *Proteins*. 2005;61(3):535–544.

334. Illergård K, Ardell DH, Elofsson A. Structure is three to ten times more conserved than sequence--a study of structural response in protein cores. *Proteins*. 2009;77(3):499–508.

335. Thangudu RR, Manoharan M, Srinivasan N, et al. Analysis on conservation of disulphide bonds and their structural features in homologous protein domain families. *BMC Structural Biology*. 2008;8:55.

336. Boisbouvier J, Blackledge M, Sollier A, Marion D. Simultaneous determination of disulphide bridge topology and three-dimensional structure using ambiguous intersulphur distance restraints: possibilities and. *Journal of biomolecular NMR*. 2000;16(3):197–208.

337. Lobstein J, Emrich C, Jeans C, et al. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories*. 2012;11(1):56.

338. Preston RJS, Rawley O, Gleeson EM, O'Donnell JS. Elucidating the role of carbohydrate determinants in regulating hemostasis: insights and opportunities. *Blood*. 2013;121(19):3801–10.

339. Allen S, Goodeve a C, Peake IR, Daly ME. Endoplasmic reticulum retention and prolonged association of a von Willebrand's disease-causing von Willebrand factor variant with Erp57 and calnexin. *Biochemical and biophysical research communications*. 2001;280(2):448–53.

340. Bahou WF, Ginsburg D, Sikkink R, Litwiller R, Fass DN. A monoclonal antibody to von Willebrand factor (VWF) inhibits factor vil binding localization of its antigenic determinant to a nonadecapeptide at the amino terminus of the mature VWF polypeptide. *Journal of Clinical Investigation*. 1989;84(July):56–61.

341. Jorieux S, Fressinaud E, Goudemand J, et al. Conformational changes in the D' domain of von Willebrand factor induced by Cys25 and Cys95 mutations lead to factor VIII binding defect and multimeric impairment. *Blood*. 2000;95(10):3796–3803.

342. Cavanagh J, Fairbrother W, Palmer III A, Rance M, Skelton N. Protein NMR spectroscopy. principles and practice. Academic Press; 2007.

343. Ulrich EL, Akutsu H, Doreleijers JF, et al. BioMagResBank. *Nucleic acids research*. 2008;36(Database issue):D402–8.

344. Berjanskii M V, Wishart DS. A simple method to predict protein flexibility using secondary chemical shifts. *Journal of the American Chemical Society*. 2005;127(43):14970–1.

345. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from nmr chemical shifts. *Journal of biomolecular NMR*. 2009;44(4):213–23.

346. Berjanskii M V, Wishart DS. Application of the random coil index to studying protein flexibility. *Journal of biomolecular NMR*. 2008;40(1):31–48.

347. Sharma D, Rajarathnam K. 13C NMR chemical shifts can predict disulfide bond formation. *Journal of biomolecular NMR*. 2000;18(2):165–71.

348. Driscoll PC, Gronenborn a M, 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 letters*. 1989;243(2):223–33.

349. Neri D, Szyperski T, Otting G, Senn H, Wuethrich 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):7510–7516.

350. Creighton TE. Disulphide bonds and protein stability. 1988;8(2):57–63.

351. Mobli M, King GF. NMR methods for determining disulfide-bond connectivities. *Toxicon*. 2010;56(6):849–54.

352. Hazes B, Dijkstra BW. Model building of disulfide bonds in proteins with known three-dimensional structure. *Protein engineering*. 1988;2(2):119–25.

353. Harrison PM, Sternberg MJ. The disulphide beta-cross: from cystine geometry and clustering to classification of small disulphide-rich protein folds. *Journal of molecular biology*. 1996;264(3):603–623.

354. Srinivasan N, Sowdhamini R, Ramakrishnan C, Balaram P. Conformations of disulfide bridges in proteins. *International journal of peptide and protein research*. 1990;36(2):147–55.

355. Schmidt B, Ho L, Hogg P. Allosteric disulfide bonds. *Biochemistry*. 2006;45:7429–7433.

356. Klaus W, Broger C, Gerber P, Senn H. Determination of the disulphide bonding pattern in proteins by local and global analysis of nuclear magnetic resonance data: application to flavoridin. *Journal of molecular biology*. 1993;232:897–906.

357. Schmidt B, Hogg PJ. Search for allosteric disulfide bonds in NMR structures. *BMC structural biology*. 2007;7:49.

358. Hogg PJ. Contribution of allosteric disulfide bonds to regulation of hemostasis. *Journal of thrombosis and haemostasis JTH*. 2009;7 Suppl 1(Table 1):13–16.

359. Bax A, Grishaev A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Current opinion in structural biology*. 2005;15(5):563–70.

360. Lipsitz RS, Tjandra N. Residual dipolar couplings in NMR structure analysis. *Annual review of biophysics and biomolecular structure*. 2004;33:387–413.

361. Chen K, Tjandra N. The use of residual dipolar coupling in studying proteins by NMR. 2012;(September 2011):47–67.

362. Bax a, Kontaxis G, Tjandra N. Dipolar couplings in macromolecular structure determination. *Methods in enzymology*. 2001;339(1997):127–74.

363. Dosset P, Hus JC, Marion D, Blackledge M. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *Journal of biomolecular NMR*. 2001;20(3):223–31.

364. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program dyana. *Journal of molecular biology*. 1997;273(1):283–98.

365. Nishikawa K, Ooi T. Radial locations of amino acid residues in a globular protein: correlation with the sequence. *Journal of biochemistry*. 1986;100(4):1043–7.

366. Laskowski R a, Luscombe NM, Swindells MB, Thornton JM. Protein clefts in molecular recognition and function. *Protein science*. 1996;5(12):2438–2452.

367. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and ProCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR*. 1996;8(4):477–86.

368. Lu W, Apostol I, Qasim MA, et al. Binding of amino acid side-chains to S1 cavities of serine proteases. *Journal of molecular biology*. 1997;266(2):441–461.

369. Otlewski J, Jelen F, Zakrzewska M, Oleksy A. The many faces of protease-protein inhibitor interaction. *The EMBO journal*. 2005;24(7):1303–10.

370. Zhang J-L, Patterson LJ, Qiu L-Y, et al. Binding between Crossveinless-2 and chordin von Willebrand factor type C domains promotes BMP signaling by blocking chordin activity. *PloS one*. 2010;5(9):e12846.

371. Pascual-García A, Abia D, Méndez R, Nido GS, Bastolla U. Quantifying the evolutionary divergence of protein structures: the role of function change and function conservation. *Proteins*. 2010;78(1):181–196.

372. Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure. *Nature reviews. Molecular cell biology*. 2007;8(12):995–1005.

373. Cook KM, Hogg PJ. Post-translational control of protein function by disulfide bond cleavage. *Antioxidants & redox signaling*. 2013;18(15):1988–2007.

374. Diao J, Maniotis AJ, Folberg R, Tajkhorshid E. Interplay of mechanical and binding properties of fibronectin type i. *Theoretical Chemistry Accounts*. 2010;125(3-6):397–405.

375. Shakhnovich BE, Dokholyan N V, DeLisi C, Shakhnovich EI. Functional fingerprints of folds: evidence for correlated structure–function evolution. *Journal of Molecular Biology*. 2003;326(1):1–9.

376. Markwick PRL, Malliavin T, Nilges M. Structural biology by NMR: structure, dynamics, and interactions. *PLoS computational biology*. 2008;4(9):e1000168.

377. Palmer AG. Nmr characterization of the dynamics of biomacromolecules. *Chemical reviews*. 2004;104(8):3623–40.

378. Boehr DD, Dyson HJ, Wright PE. An NMR perspective on enzyme dynamics. *Chemical reviews*. 2006;106(8):3055–79.

379. Kay L. Protein dynamics from NMR. *Biochem Cell Biol*. 1998;76(2-3):145–152.

380. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. *Nature*. 2007;450(7172):964–72.

381. Blackledge M. Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2005;46(1):23–61.

382. Bouvignies G, Bernadó P, Meier S, et al. Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proceedings of the National Academy of Sciences*. 2005;102(39):13885–13890.

383. Lindorff-Larsen K, Best RB, Depristo M a, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. *Nature*. 2005;433(7022):128–32.

384. Bouvignies G, Vallurupalli P, Hansen DF, et al. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature*. 2011;477(7362):111–114.

385. Baldwin AJ, Kay LE. NMR spectroscopy brings invisible protein states into focus. *Nature chemical biology*. 2009;5(11):808–14.

386. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in biochemical sciences*. 2010;35(10):539–46.

387. Osawa M, Takeuchi K, Ueda T, Nishida N, Shimada I. Functional dynamics of proteins revealed by solution nmr. *Current opinion in structural biology*. 2012;22(5):660–9.

388. Casonato A, Sartorello F, Cattini MG, et al. An arg760cys mutation in the consensus sequence of the von willebrand factor propeptide cleavage site is

responsible for a new von willebrand disease variant. *Blood*. 2003;101(1):151–156.

389. James PD, Lillicrap D. The molecular characterization of von willebrand disease: good in parts. *British journal of haematology*. 2013;(February):166–176.

390. Allen S, Abuzenadah a M, Blagg JL, et al. Two novel type 2N von Willebrand disease-causing mutations that result in defective factor VIII binding, multimerization, and secretion of von Willebrand factor. *Blood*. 2000;95(6):2000–2007.

391. Schneppenheim R, Lenk H, Obser T, et al. Recombinant expression of mutations causing von Willebrand disease type Normandy: characterization of a combined defect of factor VIII binding and multimerization. *Thrombosis and haemostasis*. 2004;92(1):36–41.

392. Castaman G, Giacomelli SH, Jacobi P, et al. Homozygous type 2N R854W von Willebrand factor is poorly secreted and causes a severe von Willebrand disease phenotype. *Journal of thrombosis and haemostasis JTH*. 2010;8(9):2011–2016.

393. Pontius J, Richelle J, Wodak SJ. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *Journal of Molecular Biology*. 1996;264(1):121–136.

394. Nilsson IM. The history of von willebrand disease. *Haemophilia*. 1999;5(Suppl 2):7–11.

395. Von Willebrand EA. Hereditary pseudohaemophilia. *Haemophilia*. 1999;5(3):223–231.

396. Singh I, Themistou E, Porcar L, Neelamegham S. Fluid shear induces conformation change in human blood protein von Willebrand factor in solution. *Biophysical journal*. 2009;96(6):2313–20.

397. Xu AJ, Springer T a. Calcium stabilizes the von Willebrand factor A2 domain by promoting refolding. *Proceedings of the National Academy of Sciences*. 2012;109(10):3742–3747.

398. Luken BM, Winn LYN, Emsley J, Lane DA, Crawley JTB. The importance of vicinal cysteines, C1669 and C1670, for von willebrand factor A2 domain function. *Blood*. 2010;115(23):4910–4913.

399. Punta M, Coggill PC, Eberhardt RY, et al. The Pfam protein families database. *Nucleic acids research*. 2012;40(Database issue):D290–301.

400. Lo Conte L, Ailey B, Hubbard TJ, et al. SCOP: a structural classification of proteins database. *Nucleic acids research*. 2000;28(1):257–9.

401. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*. 2004;32(5):1792–7.

402. Waterhouse AM, Procter JB, Martin DM a, Clamp M, Barton GJ. Jalview version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189–91.

403. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22(12):2577–637.

404. Eswar N, Webb B, Marti-Renom MA, et al. Comparative protein structure modeling using Modeller. *Current protocols in protein science*. 2007;50(2.9):1–31.

405. Delaglio F, Grzesiek S, Vuister GW, et al. NMRPipe: a multidimensional spectral processing system based on Unix pipes. *Journal of biomolecular NMR*. 1995;6(3):277–93.

406. Hansen DF, Yang D, Feng H, 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. *Journal of the American Chemical Society*. 2007;129(37):11468–79.

407. Fan QR, Hendrickson W a. Structure of human follicle-stimulating hormone in complex with its receptor. *Nature*. 2005;433(7023):269–277.