

Jeffrey J. Sutherland · Lee A. O'Brien ·
David Lillicrap · Donald F. Weaver

Molecular modeling of the von Willebrand factor A2 Domain and the effects of associated type 2A von Willebrand disease mutations

Received: 8 March 2004 / Accepted: 6 May 2004 / Published online: 3 August 2004
© Springer-Verlag 2004

Abstract A homology model for the A2 domain of von Willebrand factor (VWF) is presented. A large number of target–template alignments were combined into a consensus alignment and used for constructing the model from the structures of six template proteins. Molecular dynamics simulation was used to study the structural and dynamic effects of eight mutations introduced into the model, all associated with type 2A von Willebrand disease. It was found that the group I mutations G1505R, L1540P and S1506L cause significant deviations over multiple regions of the protein, coupled to significant thermal fluctuations for G1505R and L1540P. This suggests that protein instability may be responsible for their intracellular retention. The group II mutations R1597W, E1638K and G1505E caused single loop displacements near the physiologic VWF proteolysis site between Y1605–M1606. These modest structural changes may affect interactions between VWF and the ADAMTS13 protease. The group II mutations I1628T and L1503Q caused no significant structural change in the protein, suggesting that inclusion of the protease in this model is necessary for understanding their effect.

Keywords Von Willebrand factor · Type 2A von Willebrand disease · A2 domain · Homology modeling · Molecular dynamics simulation

Electronic Supplementary Material Supplementary material is available in the online version of this article at <http://dx.doi.org/10.1007/s00894-004-0194-9>

Introduction

Von Willebrand factor (VWF), a large protein found in blood, plays several essential roles in the early stages of blood-clot formation. The functions of VWF are attributed to several of its domains which are found in the following arrangement: D1–D2–D'–D3–A1–A2–A3–B1–B2–B3–C1–C2; [1] the mature protein (lacking the D1 and D2 propolypeptides) consists of 2,050 residues. When a blood vessel is damaged, the underlying cells found in the subendothelium are exposed to blood. VWF is capable of binding to the subendothelial protein collagen via its A1 and A3 domains and this causes a conformational change in VWF, thus exposing its platelet-binding sites in the A1 and C1 domains. [2, 3] VWF can bind to several platelets and these platelets aggregate forming a platelet plug, which will eventually lead to blood-clot formation. VWF undergoes extensive post-translational modifications including dimerization through multiple intramolecular disulfide bonds between the carboxyl-terminal ends of the protein and then, once transported to the Golgi, multimerization through interdimer disulfide bonds between the amino-terminal ends. [4, 5] The resulting multimers range in size from about 500 to 20,000 kDa and it is the highest-molecular weight multimers that are most functionally active. [6, 7] Regulating multimer size is imperative for proper hemostasis, because circulation of ultra-large VWF multimers results in the disorder thrombotic thrombocytopenic purpura (TTP) and a reduction in the size of multimers results in the bleeding disorder known as von Willebrand disease (VWD). [8, 9]

J. J. Sutherland
Department of Chemistry,
Queen's University,
Kingston, Ontario, K7L 3N6, Canada

L. A. O'Brien · D. Lillicrap
Department of Pathology,
Queen's University,
Kingston, Ontario, K7L 3N6, Canada

D. F. Weaver (✉)
Departments of Medicine (Neurology) and Chemistry
and School of Biomedical Engineering,
Dalhousie University,
Halifax, Nova Scotia, B3H 4J3, Canada
e-mail: weaver@chem3.chem.dal.ca
Tel.: +1 902-494-7183
Fax: +1 902-494-1310

The size of VWF is regulated by the recently identified plasma protease ADAMTS13, which cleaves VWF between Tyr¹⁶⁰⁵ and Met¹⁶⁰⁶ in the A2 domain (i.e. the proteolysis site). [10, 11] The A2 domain consists of residues 1,480–1,672 of VWF. Mutations in the A2 domain may result in Type 2A VWD which is characterized by decreased platelet-dependent function resulting from the absence of large and intermediate molecular weight VWF multimers. [12, 13, 14] Two distinct pathogenic mechanisms cause Type 2A VWD. [12, 13, 14, 15] Group I mutations are characterized by impaired intracellular transport, storage and secretion of high molecular weight multimers, whereas Group II mutations are characterized by increased susceptibility of the VWF protein to proteolysis by the ADAMTS13 protease. It is currently not known why some mutations in the A2 domain result in group I mutations whereas others result in group II mutations.

When an atomic model of a protein is available, the changes induced by a mutation can be studied with molecular dynamics simulation. However, there is no experimentally determined 3D structure for the VWF A2 domain. Homology modeling is used to predict the structure of proteins from related proteins of known structure. [16] Jenkins et al. have used homology modeling to predict the structures of the A1, A2 and A3 domains of VWF. [17] Since the publication of their work, the structures of the A1 and A3 domains, and the related integrins $\alpha 1-\beta 1$ and $\alpha 2-\beta 1$, have been determined by X-ray crystallography. We have developed a homology model of the VWF A2 domain using six proteins of known structure sharing from 22–25% sequence identity with the A2 domain, and have used it to study several group I and II type 2A VWD mutations associated with the A2 domain of VWF.

Methods

Template selection, template alignment, target-template alignment

Templates related to the A2 domain (the target) were identified by searching the protein data bank (PDB) using BLASTP. [18] We have used the CE program [19] for obtaining a multiple sequence alignment of templates from their structural superposition. To obtain a multiple alignment, templates were added by progressively merging those pairwise alignments that contained the fewest gaps or insertions. The final structural superposition was refined with the program QUANTA97 (Accelrys Inc.; San Diego, CA) by superimposing only those residues in conserved regions.

In this work, we used target–template alignments produced by a variety of methods that can be grouped into four categories. The term “profile” is used to describe a scoring (substitution) matrix in which the score associated with a substitution is modified to reflect its observed frequency at the corresponding position in related se-

quences. In contrast, generic scoring matrices (e.g. the BLOSUM matrices) use the same score for a substitution irrespective of its position in the sequence.

- (a) Profile–sequence: using the target sequence, three PSI-BLAST [18] iterations were performed with the non-redundant database and an E -value cutoff of 1×10^{-4} for inclusion in the profile derivation. The profile was saved and used to align the target to each template structure using the Smith–Waterman local alignment algorithm [20] implemented in the program BLASTPGP. The alignment was repeated using profiles constructed with each of the templates (12 alignments). The web servers 3D-PSSM, [21] bioinbgu (SEQPRF, PRFSEQ) [22] and mGenTHREADER [23] generated 15 alignments.
- (b) Profile–profile: the PSI-BLAST profiles for the target (sequence length n) and templates (length m) were saved in text form. An $n \times m$ matrix was formed by calculating the matrix product of the target profile ($n \times 20$) and the transpose of each template profile ($20 \times m$). Gap penalties are arbitrary quantities that must be fixed through benchmarking. In order to transfer the default values used by PSI-BLAST in profile–sequence alignment, the scores in the $n \times m$ matrices were normalized such that they have the same mean and standard deviation as the individual PSI-BLAST profiles of the target and templates. The optimal path through the normalized matrix was searched using the Needleman–Wunsch global alignment algorithm [24] implemented in the program MODELLER [25] (six alignments). The web servers FFAS [26] and 3D-Jigsaw [27, 28] generated seven alignments.
- (c) Hidden Markov models: the web servers SAM-T99 [29] and HMMER [30] seeded with a PFAM database alignment [31] generated 11 alignments.
- (d) Multiple alignments using generic substitution matrices: using the ALIGN2D routine in MODELLER, the block of template structures was aligned with a block of four high-identity sequences (64% to 75% identity with the target), identified with PSI-BLAST. The web server bioinbgu (GONPM) generated one alignment.

The 53 alignments were merged together by using the CE-derived structural alignment of the templates as an anchor. A small number of consensus-derived target–template alignments were sought for model building. The following criteria were applied to eliminate alignment variants from further consideration: (i) insertions or deletions in conserved regions of the templates, (ii) failure to reproduce conserved salt-bridges and H-bonds in the alignment (there are no disulfide bonds), (iii) alignments in loop regions where the length of a template differs by two or more amino acids from the length of the target, (iv) physically impossible alignments as deduced from the known structures (i.e. stretching too few amino acids over a distance).

Model building

The program MODELLER was used to create homology models. Because PDB structures 1aox and 1qc5 are highly similar, only the latter was retained. A thorough optimization procedure was used to generate 20 models from 20 randomized starting configurations (non-default parameters: LIBRARY_SCHEDULE=1, MAX_VAR_ITERATIONS=300; MD_LEVEL='refine_4'; REPEAT_OPTIMIZATION=3, MAX_MOLPDF=1E6).

A clustering procedure was used to produce a consensus model (see TRANSFER_XYZ in the manual). The ten best models from 20 runs were superposed using the MALIGN3D command. For each protein atom, the equivalent atoms from the ten models were clustered and the average coordinates taken from the largest cluster. The resulting clustered model was relaxed using conjugate-gradient minimization of the probability density function (PDF). The cluster size was varied from 0.3 Å to 2.0 Å, and the model with the lowest PDF was retained. In all cases, this led to a model with a lower value of the PDF than the best of 20 single models.

Models containing various mutations were created by substituting the appropriate amino acid in the wild-type (WT) model using QUANTA.

Molecular dynamics simulation

Using the program SOLVATE, [32] the protein was surrounded by a large solvent sphere containing NaCl electrolyte at physiological concentration, with ions placed according to a Debye–Hückel distribution. This was truncated to a rectangular parallelepiped, with solvent extending at least 10 Å in every direction from the protein. The side chains of Asp, Glu, Lys, Arg and His were set to their charged state. It was decided to charge the two His residues because of the large negative charge (−8) of the protein. For the WT protein, this gave an overall neutral system (SOLVATE added more Na⁺ than Cl[−]). For mutations that affected the charge of the protein, one ion located far from the protein was removed to neutralize the system. The WT system consisted of 20,883 atoms (one protein, 6,040 waters, 32 ions).

Molecular dynamics simulations were carried out using the program NAMD2, [33] with the CHARMM27 all-atom force field. [34] Accordingly, waters were represented by the TIP3P model. [35] Bonds to all hydrogen atoms were kept rigid using SHAKE, [36] permitting a time step of 2 femtoseconds (fs). The system was simulated in periodic boundary conditions, with full electrostatics computed using the particle mesh Ewald (PME) method [37] with a grid spacing on the order of 1 Å or less. A cutoff distance of 10.0 Å was used for van der Waals interactions, with a switching function from 8.5 Å to 10.0 Å, and a patch size of 12.0 Å (i.e. the pairlistdist parameter). With the protein atoms fixed, the solvent was subjected to 996 steps of conjugate-gradient energy minimization, followed by 50 picoseconds (ps) of equil-

ibration. With only the protein backbone atoms fixed, the entire system was subjected to 480 minimization steps. The backbone constraints were then removed, but C α atoms were subjected to a harmonic restraining potential of 25 kcal mol^{−1} Å² and the system subjected to a further 480 minimization steps. The system was equilibrated for 1,000 ps; the harmonic restraints were removed at 150 ps. For equilibration, the NpT ensemble was used (constant pressure and temperature). Pressure was maintained at 1 atm using the Langevin piston method, [38] with a piston period of 200 fs, a damping time constant of 500 fs, and piston temperature of 310 K. Temperature was controlled with Langevin dynamics, using a coupling constant of 5 ps^{−1}. After equilibration, temperature and pressure control were disabled (NVE ensemble). The system was simulated for a further 9 ns. Coordinates were saved every 1 ps after the removal of harmonic restraints. During the equilibration phase, the r-RESPA integrator [39] was used so that non-bonded interactions could be calculated at every three steps. This was disabled after equilibration, as the system energy drifted unacceptably. Each 10-ns simulation takes 15 days on a SGI Onyx 3800 with 16 R14K 400-MHz processors. The trajectories were analyzed using QUANTA and InsightII 2000 (Accelrys Inc.; San Diego, CA). Solvent accessibility and secondary structure were evaluated using the algorithms of Lee and Richards, [40] and Kabsch and Sander, [41] as implemented in QUANTA.

Results

Template selection, template alignment, target–template alignment

Six template proteins were identified by searching the PDB with BLASTP (Table 1). Other sequence search and fold-recognition methods identified the same set of templates. All templates adopt a flavodoxin-like dinucleotide binding (Rossmann) fold, [42] with a core of five parallel β -sheets and one anti-parallel β -sheet surrounded by six α -helices.

When the structures of proteins are available, the correct alignment of sequences can be deduced from their structures. Excluding the high-identity 1qc5–1aox pair, the pairwise sequence identity of the templates ranged from 16.1% to 34.8%, with C α root mean square deviations (RMSD) of 1.3 Å to 2.1 Å following superposition of structures with CE. [19]

The most critical step in homology modeling is the target–template alignment. For sequences having roughly 25% identity (Table 1), obtaining the correct alignment is difficult. [43] A consensus alignment was obtained by merging 53 alignments from an array of methods. Following this process, only a few distinct alignments remained (Fig. 1). The variants “int” and “A” segregate with the template used in the alignment (i.e. an integrin or an A domain, respectively). The variant “shift” is obtained by shifting the C-terminal end of the target by one

Table 1 Description of templates used for homology modeling

Template description	PDB code ^a	Resolution (Å)	% identity with target ^b	Alternative templates
VWF A1 domain	1auq ^c	2.3	22	1oak, 1fns
VWF A3 domain	1atz_A	1.8	24	1ao3, 1fe8
integrin $\alpha 1$ - $\beta 1$	1qc5_A	2.0	23	1ck4
integrin $\alpha 2$ - $\beta 1$ ^e	1aox_A ^d	2.1	22	1dzi
integrin CD-11a ^c	1zon	2.0	21	1cqp, 1zop, 1zoo, 1lfa, 1dgq
integrin CD-11b ^c	1jlm	2.0	25	1ido, 1idn, 1bhq, 1bho

^a pdb_A indicates that chain A is used

^b Sequence identity as derived from final target-template alignment (see Methods)

^c Missing surface side chain: K660 was copied from 1fns_A

^d Missing surface side chains: K168, E204, R243 set to most probable Dunbrack and Karplus rotamer in QUANTA.

^e Unoccupied ion coordination site

residue in the N-terminal direction. The alignment variants “int”, “A” and “shift” give a good agreement between the predicted [44, 45, 46] and observed secondary structures. Models were built using these three alignments, since evaluating an alignment as embodied in a model is often more reliable than using sequence information alone.

Our alignments differ from that of Jenkins et al. [17] in a key area: the segment from position 99 to 138 (Fig. 1) is shifted by three residues towards the C-terminus in their alignment. In the consensus alignment, there is near unanimity close to the proteolysis site, with only two of 53 methods producing alignments shifted by one residue towards the C-terminus. Their alignment also appears problematic when comparing predicted secondary structure to that of the templates. Nonetheless, we evaluated this alignment further by constructing an additional model because of the importance of this region in deriving conclusions from the model.

Model building

The program MODELLER generates a protein model through the satisfaction of spatial restraints. The restraints, expressed as conditional probability density functions (PDFs), are obtained by assuming the corresponding distances and angles between aligned residues in the templates and the target structure are similar. In addition to information from target–template alignments, additional restraints are obtained from CHARMM force-field parameters [34] for bond-lengths, bond angles, and non-bonded atom contacts. The individual PDFs are combined into a global function; a structure that best satisfies this function is sought using numerical optimization.

Aligning target residues with template residues instructs the program to derive restraints from those template residues. For the conserved sheets and helices, all five templates were used for the corresponding residues in the target. The models differ in the templates selected for loop regions. For the models *m_int* and *m_A*, all templates having no gaps with respect to the target were used for building loops. The model *s* uses only the “best” template

structure at each loop, as determined by a consensus of BLOSUM62, PSI-BLAST profile–sequence, and profile–profile scores for the loop (the labels “m” and “s” indicate multiple or single template for loops). When other template loops had similar C α structure as the highest scoring loop, they were used for supplying additional side-chain restraints in the model *s*. For each of these, the C-terminal shift in sequence was applied (see above). This gave a total of seven models: *m_int*, *m_A*, *s*, *m_int_shift*, *m_A_shift*, *s_shift* and *jenkins*.

The seven models were evaluated with the structure validation tool ProsaII. [47] The sum of surface and C β pair statistical potentials is evaluated at each amino acid in the structure. Negative values of the sum of potentials are indicative of a well-folded protein. It was found that the models using the “shift” alignment have a substantially higher potential at the C-terminal end than the other models, and that *jenkins* has a higher potential over the residues 1,578–1,627 (Fig. 2). Those models were discarded. The comparison of potential profiles for the templates with those of the models *m_int*, *m_A* and *s* provides some indication they are reasonable. However, for the regions L6 (residues 1,591–1,600) and L10 (residues 1,642–1,648) the potentials for the models lie significantly above those of the templates (see Fig. 1 for the definition of regions). The latter also exhibits the largest violations of the MODELLER restraints. This suggests that the models are less reliable over these regions.

In order to study the structural effects of mutations on the A2 domain, it was necessary to select a single model for performing molecular dynamics simulation. Model *m_int* was selected over *m_A* since the latter has larger violations of the restraints at alignment positions 58–66. Model *s* was selected over *m_int* as it resulted in smaller C α RMSD from the model at the end of the 1-ns equilibration (1.6 Å and 2.2 Å, for *s* and *m_int*, respectively). The difference was largely confined to loop L7, indicating that it may be more accurately modeled in *s*. Henceforth, all references to the homology model concern *s*.

The location of key features are mapped onto the model in Fig. 3. These include the proteolysis site, two glycosylation sites Arg¹⁵¹⁵ and Arg¹⁵⁷⁴, and the mutations listed in Table 2. The mutations are not located within

Fig. 1 Final consensus alignment obtained from multiple methods, showing the sequence of A2 residues 1,496–1,669 that could be modeled. Alignment “A” is shown below “int” at positions 58–66. Alignment “shift” is shown below at positions 176–202. The residue numbers for the A2 domain are indicated as “A2 seq” (gaps are not numbered). For secondary structure, “E” indicates β -sheet and “H” indicates α -helix. Only the secondary structure prediction for Psipred [44] is shown; PHD [45] and Jpred [46] gave similar results. The secondary structure is the same for all templates, except for positions 135–139 where that for 1zon is shown. The secondary structure for the model is implied through the region labels B1, L1, A1, etc., with “B” corresponding to β -sheet, “A” to α -helix and “L” to loop. The model solvent accessibility is indicated by integers 0–9 where 0 indicates 0–9% accessible, 1 indicates 10–20% accessible, etc.

aln pos	10	20	30	40	50	60	
1auq	LLDLVFLLDGSSRLSEAEFEVLKAFVVDMMERLRISQKWVAVVEYHDGSHAYIGLKD--RKRPSSEL						
1atz	PLDVILLDDGSSFPASYFDEMKSPAKAFISKANIGPRLTQVSVLQYGSITTTIDVPWNV--VPEKAHL						
1aox	LIDVVVVCDESNISY--PWDVKNFLEKPVQGLDIPGPKTQVGLIQYANNPRVFNLTNTYKTKR--EM						
1qc5	QLDIVIVLDGNSIY--PWDSVTAFNLDDLLEKMDIGPKQKQVGVIVQYGENVTHEFNLNKYSSTE--EV						
1jlm	DSDIAPFLIDGSGSIIIPHDFRRMKEFVSTVMEQLKKS--KTLFSLMQYSEEFRIHFTFKFQNNP--NP						
1zon	NVDLVFLPDGSMQLQDFQKILDFMKDVMKLSNT--SYQFAAVQFSTSYKTEPFDSDYVVKR--DP						
A2	VLDVAFVLEGSDKIGEADFNRSKEFMEVVIQRMDVQDSIHVTVLQYSYVMVTVEYFPFSEAQSKG--DI						
					SE--AQSKG		
A2 seq	1505	1515	1525	1535	1545	1555	
psipred	-EEEEEEEE-----HHHHHHHHHHHHHHHHHH-----EEEEEEEE-----EEEEEE-----HH--HH						
template	-EEEEEEEE-----HHHHHHHHHHHHHHHHHH-----EEEEEEEE-----EEEEEE-----HH--HH						
region	< B1 >< L1 ><	A1	>< L2 ><	B2	><L3><B3><	L4 >< A2	
solv acc	3100000000054086820330140024004308036941000000005705320304307666--50						
aln pos	70	80	90	100	110	120	130
1auq	RRIASQVKYAGSQVASTSEVLKYTLFQIFSKI--DRPEASRIALLMASQEPQORMSRNFV--RYVQG						
1atz	LSLVDVMQREGG-PSQIGDALGFVRYLTSEMHGARGPASKAVVILVTDVSV--DS--VD---AAADA						
1aox	IVATSQTSQYGGDLTNTFGAIQYARKYAYSAAAGGRRSATKVMVVVTDGESH---D--GSMLKAVIDQ						
1qc5	LVAAKKIVQRGGRTMTALGTDTRARKEAFTEARGARRGKVKVMVITVDGESH---D--NHRLLKVIQD						
1jlm	RSLVKPITQLLG-RTHATGIRKVVRELFNITNGARKNAFKILVVIITDGEKF--GD--PLGYEDVIPE						
1zon	DALLKHVKHMLL-LTNTFGAINYVATEVFREELGARPDAATKVLIIITDGEAT--DS-----GNIDA						
A2	LQRVREIRYQGNRNTGLALRYLSDHSLVVSQGDREQAPNLVYMTGNPAS--DE-----IK-RL						
A2 seq	1565	1575	1585	1595	1605	1615	
psipred	HHHHHHHH-----HHHHHHHHHHHHHHHHHH-----EEEEEEEE-----HH--HH						
template	HHHHHHHH-----HHHHHHHHHHHHHHHHHH-----EEEEEEEE-----HH--HH						
region	><	L5 ><	A3	><	L6 ><	B4 ><	L7 ><
solv acc	2730660922148601004005201450057612018906300000001503--39-----07-43						
aln pos	140	150	160	170	180	190	200
1auq	LKKKVVIVIPVIGIGPHA-----NLKQIRLIEKQAPENKAFVLSVDELEQQORD---EIVSYLCD						
1atz	ARSNRVTVPFPIGIDRY-----DAAQLRILAGPAGDSNVVKLQRIEDLPTMVTLGNSTFLHKL--						
1aox	CNHDNILRFGIAVLGYNLNRNALDTKNLIKEIKAIASIPTEYFFNVSDAALLEKAG---TLGEQIFIS						
1qc5	CEDENIQRFSIALILGYSYRNLSTKFKVVEIKSIASEPTEKHFNFVSDIETALVTIVK---TLGERI--						
1jlm	ADREGVIRYVIGVDAF-----RSEKSRQELNLTASKPPRDHVFVQNNFPAALKTION---QLREKIFA						
1zon	AK--DIIRYIIGIGKHF-----QTKESQETLHKFASKPASEFVKILDTFEKLDKDLFT---ELQKKI--						
A2	PG--DIQVVPVIGVGNP-----NVQELERIGWP---NAPILIQDFETLPREAP---DLVLQRC-						
					NAPILIQDFETLPREAP--LVLQRC		
A2 seq	1625		1635	1645		1655	1665
psipred	-----EEEEEEEE-----HHHHHHHH-----EE--HHHHHHHHHH-----HHHH--						
template	HH-----EEEEEEEE-----HHHHHHHH-----EEE--HHHHHHHHHH-----HHHHHH--						
region	L8 ><	B5 ><	L9	><	A4 ><	L10 ><	B6 ><
solv acc	01--9040000003850-----15810660058---15315084051056005---6017306-						

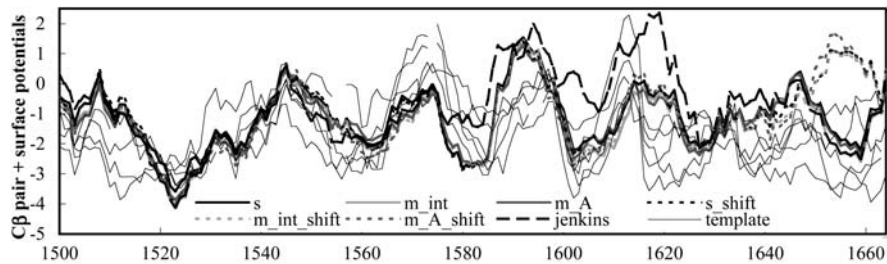


Fig. 2 Sum of ProsaII $C\beta$ pair and surface potentials as a function of residue. For clarity, the potentials of the templates (indicated as *fine gray lines*) are shown only for residues that are aligned with the A2 domain. The potentials are averaged over a ten-residue window

particular regions of the domain and correspond mostly to buried residues.

Molecular dynamics simulation of WT and mutant models

Simulations were performed to investigate the structural and dynamic effects of mutations on the WT model. In order to draw conclusions about mutation-induced changes, it is necessary to perform an equivalent simulation for the WT protein. To increase confidence in our results, we performed two simulations for the WT protein

starting from different sets of initial velocities. The all-atom RMSD with respect to the model is shown in Fig. 4 for the two WT simulations, along with those for G1505E and G1505R. For both simulations, the WT protein remained stable throughout, with no indications of unfolding. For comparing the structural effects of mutations, an average structure was calculated from snapshots taken every 1 ps during the final 3 ns of simulations. The $C\alpha$ RMSD for the time-averaged and final structures are given in Table 3. The simulations of the mutants also indicate that no dramatic unfolding occurs. In order to gain insight into local structural changes, it is useful to examine the $C\alpha$ RMSD from the model of the time-av-

Fig. 3 Schematic representation of the VWF A2 domain model (two views). The $C\alpha$ atoms of mutated residues are shown as green spheres. Residue 1606 of the proteolysis site is shown as a red sphere. Two glycosylation sites are also shown. Region labels are described in Fig. 1

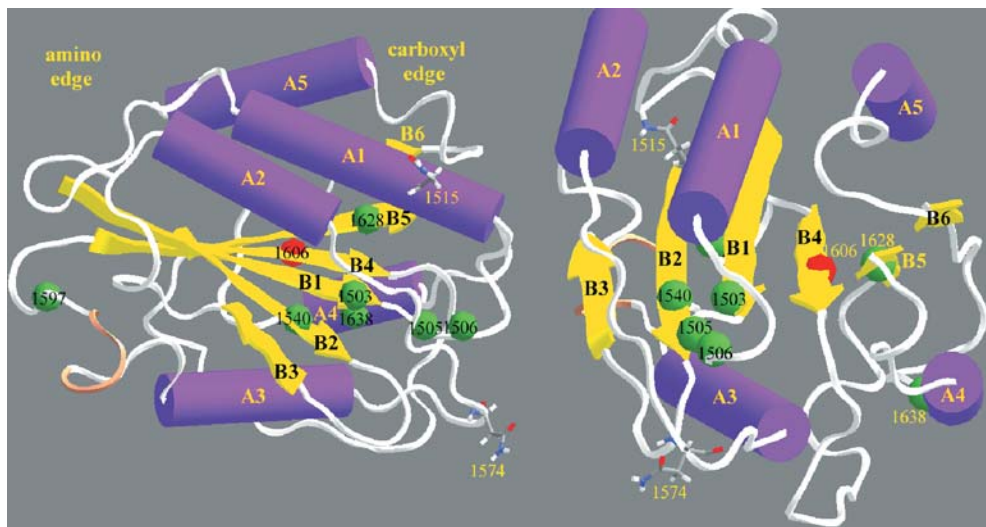


Table 2 Mutations modeled with molecular dynamics simulation

	Mutation	Residue solvent accessibility (in WT model) (%)	Comment ^a
Group II	R1597W	19	Most common Type 2A mutation, reported in patients from the US, Japan, France and England
	E1638K	15	Loss of high and intermediate molecular weight multimers
	I1628T	0	Loss of high and intermediate molecular weight multimers
	L1503Q	1	Novel mutation resulting in loss of only the highest molecular weight multimers
	G1505E	0	Loss of high and intermediate molecular weight multimers
Group I	G1505R	0	Loss of high and intermediate molecular weight multimers
	L1540P	0	Loss of high and intermediate molecular weight multimers
	S1506L	2	Loss of high and intermediate molecular weight multimers, reported in patients from the US, Japan, Argentina and Israel

^a VWF database: <http://www.shef.ac.uk/vwf/>

eraged structure as a function of residue (Fig. 5a). It is important to interpret changes associated with mutations in consideration of the deviations observed for the WT simulations and the deviations of the templates from their average structure. The latter can be taken as a measure of the local reliability of the model. Deviations for mutants that fail to exceed these quantities are not significant. Also, interpreting which deviations are significant is region-dependent; a deviation less than 2 Å in a loop region is likely to be insignificant, whereas a similar deviation in a sheet region would amount to a drastic change. In addition to structural changes, fluctuations about the average structure were examined during the final 3 ns of the simulation (Fig. 5b).

Before discussing specific mutations, some general remarks can be made. Nearly all simulations produced large deviations from the model over the residues 1,592–1,596 (loop L6) and loop L9, similar in magnitude to the deviation among the templates. The two regions are the same (L6) or adjacent to (L9) those discussed above, where high ProsaII potentials and MODELLER violations were observed. For the region L9–A4, it was observed that all simulations except those for E1638K and G1505R (see below) resulted in similar structures, with A4 shifted outwards from its position in the model. Finally, the

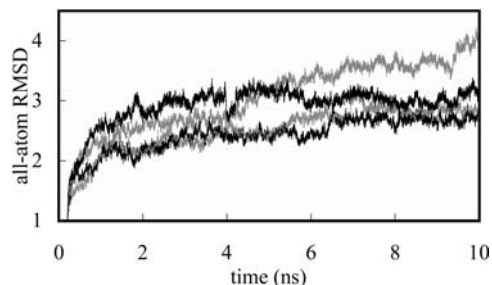


Fig. 4 All-atom root-mean-square deviation from the model as a function of time. At 10 ns, the series from bottom to top are: WT1 (black), WT2 (gray), G1505E (black), G1505R (gray)

variability over the C-terminal region A5 is attributed to the truncation of VWF at residue 1,669.

R1597W (Group II)

The mutation causes a loss of two hydrogen bonds (H-bonds) between NH_2 and NH of Arg¹⁵⁹⁷, and each carboxyl O of Asp¹⁴⁹⁸ (functional groups are for side chains unless noted otherwise). In the WT simulation, the other Arg¹⁵⁹⁷ NH_2 group forms an H-bond with the backbone

Table 3 C α root mean square deviation with respect to the homology model for time-averaged and final structures

Mutation	C α RMSD from homology model (Å)	
	Time-averaged structure	Final structure (10 ns)
WT1	1.67	2.09
WT2	1.64	1.87
R1597W	2.45	2.79
E1638K	2.14	2.18
I1628T	2.01	2.05
L1503Q	1.96	2.11
G1505E	1.97	2.28
G1505R	2.64	2.91
L1540P	2.52	3.08
S1506L	2.08	2.18

C=O of Asp¹⁵³³, located at the end of sheet B2. In the mutant, the phenyl ring of Trp¹⁵⁹⁷ is involved in edge-to-face stacking with the imidazole ring of His¹⁵³⁶ (2.0 Å separation) (Fig. 6a, b). A significant shift occurs over the residues 1,597–1,602^A (superscripted letters refer to Fig. 5). The segment is adjacent to the high RMSD residues 1,592–1,596 (L6), and deviations in 1,597–1,602 may appear related. However, the templates and WT simulations deviate little over 1,597–1,602, with residues 1,601 and 1,602 being part of sheet B4. Even over the region 1,592–1,596, R1597W is clearly distinct from the others, having a truncated five-turn at 1,594–1,595. This

appears to cause a 2-Å shift of helix A3^B towards the carboxyl edge of the protein. Of all simulations, R1597W has Pro¹⁶⁰¹ closest to Asp¹⁶²² in L8^C (Pro¹⁶⁰¹-Asp¹⁶²² distance of 2.8 Å in the mutant simulation, compared to 5.1 Å in the WT simulation). The displacement of residues 1,601–1,602 towards L8 may contribute to the latter's shift, although other mutant simulations resulted in a shift of similar magnitude without the Pro¹⁶⁰¹-Asp¹⁶²² close contact. Sheet B4 and loop L8 lie at either end of the displaced loop L7^D, located near the proteolysis site. The shifted helix A3, of which the carboxyl edge approaches loop L7, may also contribute to its displacement. Thus, it appears that a mutation in an exposed, non-core region well away from the proteolysis site effects a structural change very near to it.

E1638K (Group II)

The mutation leads to the formation of two H-bonds; the first is between NH₃ of Lys¹⁶³⁸ and a carboxyl O of Glu¹⁵⁰⁴ (N-H ... O distance of 1.6 Å) while the second is between NH₃ of Lys¹⁶³⁸ and the hydroxyl O of Thr¹⁵⁷⁸ (N-H ... O distance of 1.8 Å) (Fig. 6c, d). The first causes a decrease in H-bonding between the carboxyl O of Glu¹⁵⁰⁴ and Thr¹⁶⁰⁸, located at the junction of B4 and L7 (OH ... O distance increases from 1.8 Å in the WT sim-

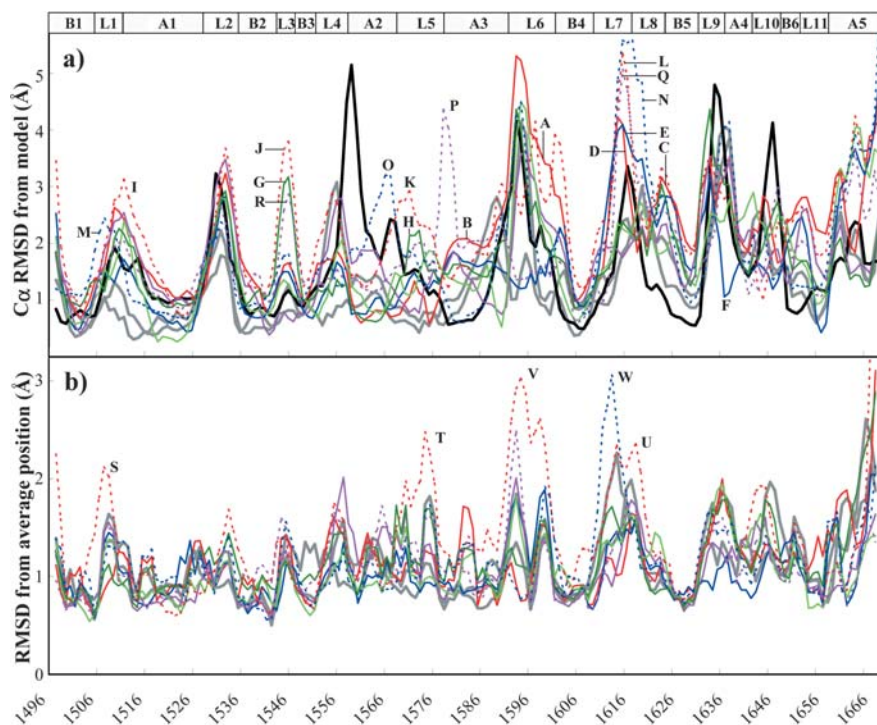
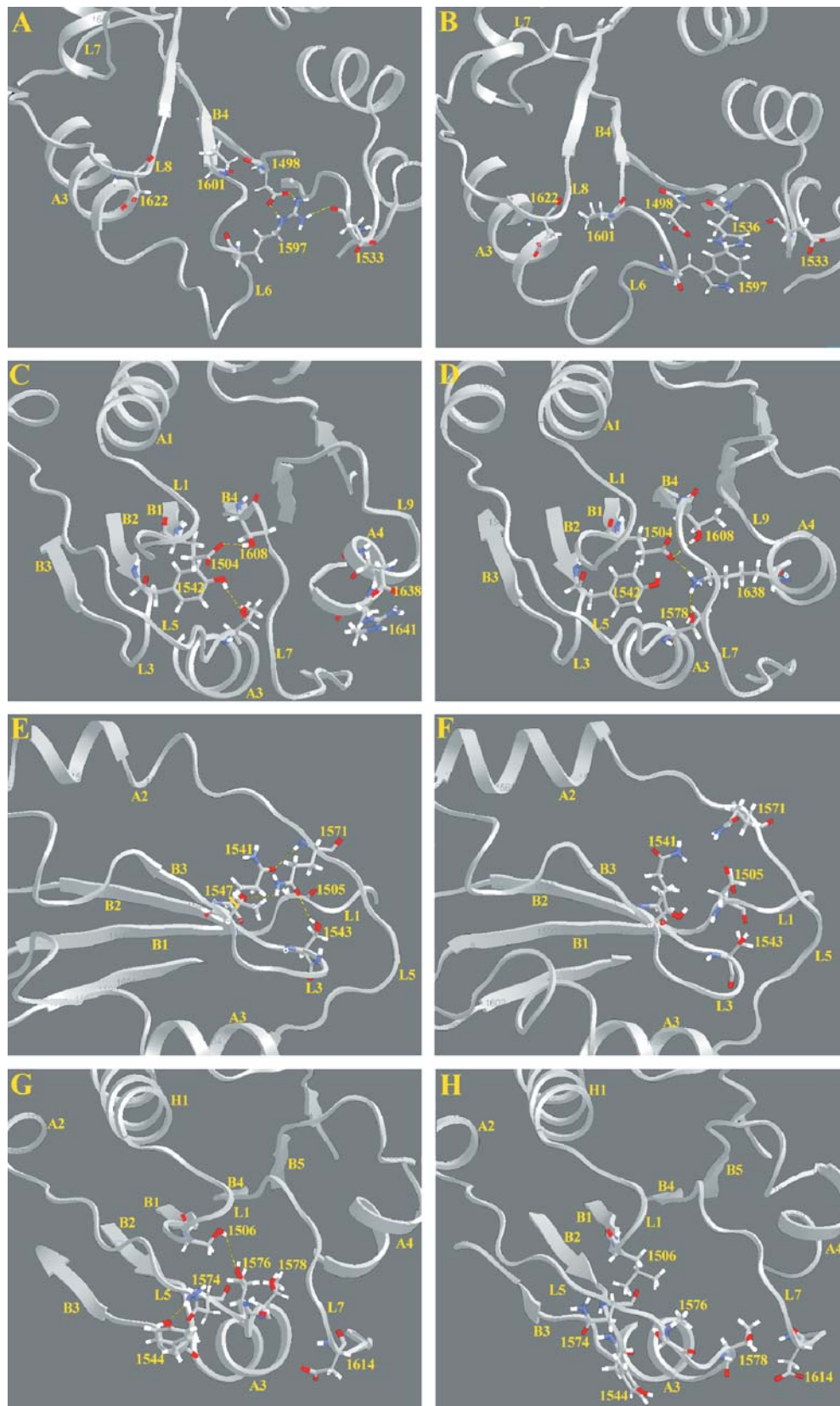


Fig. 5 Root-mean-square deviations (RMSD) as a function of residue, averaged over a three-residue window. **a** C α RMSD from the homology model of the time-averaged structures. **b** RMS displacement (fluctuations) of residues about their time-averaged position. Temperature factors familiar from X-ray crystallography are obtained by squaring the RMS fluctuation and multiplying by $8\pi^2/3$. For reference, the RMSD of the templates with respect to their

average structure is indicated in bold black in panel **a**. The WT simulations are in bold gray, group I mutants are shown with dotted lines (G1505R in red, L1540P in blue, S1506L in pink) while group II mutations are shown with solid lines (R1597W in red, E1638K in blue, I1628T in pink, L1503Q in light green, G1505E in dark green). Letter labels are discussed in the text. Region labels are described in Fig. 1

Fig. 6 Schematic representations of structures after 10 ns of simulation for WT protein (*left panels*) and mutants (*right panels*). R1597W (**a, b**), E1638K (**c, d**), G1505E (**e, f**) and S1506L (**g, h**). Region labels are described in Fig. 1. H-bonds are indicated as dotted lines. For G1505E, the L6 label is positioned between residues 1,594–1,595 (i.e. at the five-turn). Some parts of the proteins are omitted for clarity



ulation to 2.3 Å in the mutant simulation). The second causes a loss of a hydrogen bonding between Thr¹⁵⁷⁸ and Tyr¹⁵⁴². This appears to have no significant consequences, perhaps because Tyr¹⁵⁴² is solvent exposed. The shifting of loop L7^E may be attributed to decreased H-bonding between Thr¹⁶⁰⁸ and Glu¹⁵⁰⁴, although the steric bulk of the Lys side chain approaching the loop may contribute. Also, it appears that the mutation tethers A4 to the core of the protein. This is reasonable in light of the H-bonds formed by Lys¹⁶³⁸, and interestingly E1638K is the only simulation for which A4^F deviates little from the model. Because of the similar L9–A4 placement achieved by nearly all other simulations, the difference between E1638K and the other proteins is probably meaningful.

I1628T (Group II)

No significant structural shifts are evident. The mutated residue is in sheet B5, adjacent to the proteolysis site in sheet B4. It is a change from a non-polar to a polar residue in the core region of the protein, surrounded by four non-polar residues.

L1503Q (Group II)

The mutation replaces a non-polar residue with a polar residue, but the formation of an H-bond with the NH₂ group of Gln¹⁵⁴¹ (N–H ... O distance of 2.1 Å) helps compensate the energetic cost of burying a polar residue. Despite this, Gln¹⁵⁴¹ maintains an H-bond with the backbone C=O of Arg¹⁵⁶⁹ also seen in the WT simulation. As for I1628T, the change occurs in a sheet region containing the proteolysis site. There is little deviation, except for loop L1, and this is not deemed significant as it does not lie outside the range observed for templates and most simulations.

G1505E (Group II)

Deviations at loop L3^G and Gln^{1571,H} appear to be related to the side chain of Glu¹⁵⁰⁵ extending towards them. For L3, one carboxyl O of Glu¹⁵⁰⁵ is 3.6 Å from a NH₂ group of Gln¹⁵⁴¹, and the other carboxyl O of Glu¹⁵⁰⁵ is 3.7 Å from the OH group of Ser¹⁵⁴³ (Fig. 6e, f). These are not H-bonds, although they are favorable electrostatic interactions between the negatively charged Glu side chain and the pocket with positive potential (as determined with DELPHI; [48] results not shown). In the WT simulation, Gln¹⁵⁷¹ is a (weak) H-bond donor for Thr¹⁵⁴⁷ (N–H ... O distance of 3.1 Å), and its backbone NH group is a H-bond donor for Gln¹⁵⁴¹ (N–H ... O distance of 2.3 Å). Gln¹⁵⁷¹ is also a H-bond acceptor for Ser¹⁵⁴³ (N–H ... O distance of 2.2 Å). If Gln¹⁵⁷¹ remained in that position, it would clash directly with Glu¹⁵⁰⁵. It is doubtful that the outward displacement of Gln¹⁵⁷¹ is significant in itself, as the protein surface should be tolerant to such changes.

Also, the distances between Glu¹⁵⁰⁵ and Gln¹⁵⁴¹ or Ser¹⁵⁴³ are too long for steric interactions to be responsible. Rather, it is likely that the loss of three H-bonds in loop L3 causes its observed deviation.

G1505R (Group I)

Large deviations occur at loop L3 and around Gln¹⁵⁷¹ as observed for G1505E, although the latter displacement involves more residues. This leads to the disruption of the same H-bonds between residue Gln¹⁵⁷¹ and Thr¹⁵⁴¹, Ser¹⁵⁴³, Thr¹⁵⁴⁷. The contacts between Arg¹⁵⁰⁵ and Gln¹⁵⁴¹, Ser¹⁵⁴³ are closer, at 2.9 Å and 2.3 Å, respectively. In contrast to G1505E, these contacts involve unfavorable electrostatic interactions. At 10 ns, the G1505R simulation contains the largest deviations observed at almost all regions of the protein. In particular, loops L1–A1^I, L3^J, L5^K and L7^L, all in proximity to the proteolysis site, contain the largest deviations for the first three loops and deviations similar to those of S1506L and L1540P for L7. From Fig. 5a, the deviation at loop L1–A1 appears to be barely beyond the range observed for other mutant models. However, Fig. 5b shows that it exhibits distinctly larger fluctuations over that region^S, beginning in sheet B1. The deviations in L5 stretch over nine residues whereas only residue 1571 of G1505E deviates by more than 2 Å from the model. Pronounced fluctuations are also observed over this region^T, and over L7^U. While no clear distinction can be made for L6 for the time-averaged structure, this region^V clearly exhibits larger fluctuations in G1505R. In contrast to the other simulations, the protein does not seem to be equilibrated at 10 ns (Fig. 4) and has the largest Cα RMSD from the model at 10 ns.

L1540P (Group I)

The mutation in sheet B2 does not cause significant structural change within that sheet. However, Pro¹⁵⁴⁰ lacks the ability to provide an H-bond donor for pairing with the carbonyl oxygen of Phe¹⁵⁰¹ in the adjacent sheet B1, and its C=O group cannot H-bond with the backbone NH in Leu¹⁵⁰³ because of geometric constraints. This causes a small shift of sheet B1^M towards sheet B2. Given the high degree of conservation of the sheet structure among the parents, it appears that the Pro substitution disrupts it sufficiently to impart other significant changes. In particular, loop L7^N is the most greatly displaced among all the models, and exhibits the largest fluctuations^W. There is some shifting reaching nearly 4 Å of the helical region A2^O that is not seen in any other model. While the Cα RMSD at 10 ns is less than for G1505R, the all-atom RMSD is larger (not shown).

S1506L (Group I)

The mutation causes a loss of hydrogen bonding with acceptor Thr¹⁵⁷⁶ (O–H ... O distance of 3.2 Å). Furthermore, the bulk of the leucine side chain causes it to approach within 3.1 Å, 2.8 Å, and 3.4 Å of the backbone of residues 1,574–1,576 in L5 (Fig. 6g, h). It does not appear that this is responsible for the shift of the loop, because of the small deviation observed at those residues. Rather, the loss of the H-bond between Thr¹⁵⁷⁶ and Ser¹⁵⁰⁶ results in a 5.3-Å shift of the Thr OH group from the WT simulation, as it exposes itself to solvent. This causes the subsequent residues 1,577–1,581 within the region L5–A3^P to shift considerably, with a noticeable unravelling of the end of A3. In this position, Thr¹⁵⁷⁸ comes within 1.1 Å of Asp¹⁶¹⁴ in loop L7 as positioned in the WT simulation. This causes L7^Q to shift away. The shift in L5 (albeit small at Asn¹⁵⁷⁴) causes a loss of H-bonding between Tyr¹⁵⁴⁴ and Asn¹⁵⁷⁴ (N–H ... O distance of 1.9 Å in the WT simulation, 5.1 Å in the mutant simulation). This may contribute to the observed shift in neighboring loop L3^R.

Discussion

It is well established that obtaining good quality homology models is difficult for targets having limited sequence similarity with the templates. [16] However, accuracy can be significantly increased when multiple sequences and structures are used, with alignments edited by hand. [49] In light of this, we have taken unusual care in constructing the target–template alignment used for building the A2 model. Where alternative alignments could not be eliminated on the basis of sequence information alone, we constructed the corresponding models in order to allow structural validation of the alignments. Statistical potential-based methods such as ProsaII [47] are known to be accurate for ubiquitous folds in the PDB, such as the Rossman fold to which A domains belong. [50, 51] We identified two regions (residues 1,592–1,596 of L6 and L9–A4–L10) over which the model is likely to be inaccurate. For the latter, it was encouraging to see a convergence for all but two molecular dynamics simulations. This suggests that that region may be better represented after refinement with molecular dynamics simulation.

The alignment of Jenkins et al. [17] differs substantially from the present alignment in the key region of the physiological proteolysis site. The Jenkins alignment places the proteolysis site within loop L7, while the present alignment places it within β -sheet B4. We found no theoretical justification for the Jenkins alignment. Not a single alignment method proposed it, the predicted secondary structure of the target and that of the templates disagreed over the highly conserved β -sheet B4, and the model built from this alignment showed significantly higher statistical potentials in the affected region. Simply examining the alignment reveals some difficulties. The residues in β -sheet B4 for the templates are predominantly non-polar. The Jenkins alignment places a Glu, Pro

and Asp residue within the sheet, residues that are the three least frequently observed in β -sheets. [52]

It is instructive to examine the experimental facts regarding proteolysis of the A2 domain by the ADAMTS13 protease. Relying on intuition, it may be thought that the proteolysis site must be located at a solvent-exposed region of the A2 domain in order for it to be accessible to the protease. Only under shear-stress conditions that cause unfolding can the protease cleave VWF in normal individuals. The situation is similar under assay conditions; VWF must be subjected to denaturants for proteolysis to be observed. [53] Because of the multi-domain structure of VWF, it may be proposed that proteolysis under shear-stress or denaturing conditions arises from inter-domain reorganization, not intra-domain structural change (i.e. rearranging beads on a string without affecting the nature of the beads). As the present study addresses only the isolated A2 domain, the burial of the proteolysis site within our model remains problematic if proteolysis only involves inter-domain changes. Recent experiments have shown that even for the isolated A2 domain, denaturing conditions are required for proteolysis to be observed. [54] This suggests that the proteolysis site is in fact not accessible to the protease under normal conditions for the isolated A2 domain.

A homology model by itself could potentially shed light on the molecular explanation of observed phenotypes, as seen for type 2M and 2B mutations of the VWF-A1 domain. [17, 51] This is not observed for type 2A mutations related to the A2 domain model presented here, with mutations not confined to a particular region of the protein and not always involving buried residues. By itself, the model provides only a limited interpretation of the effect of mutations.

When the 3D structure of a protein is known, molecular dynamics simulation can be used to assess the structural change caused by mutations. For group II mutations, we observed either local changes in one loop, or no change at all. We speculate that the displacement of loop L7 in R1597W and E1638K, and that of L3 in G1505E, which are located in proximity to the proteolysis site, lead to greater susceptibility to proteolytic attack. For I1628T and L1503Q, we observed no significant structural changes. Both occur in proximity to the proteolysis site; the observed phenotype may result from specific interactions of these residues with the protease, a scenario that cannot be evaluated with the present model. For the group II mutations studied here, there is no evidence of structural instability that might prevent proper folding of the domain. For group I mutations, the changes that were observed were less subtle, involving larger deviations over many regions of the protein. The mutation G1505R produced large deviations from the model in loops L1, L3, L5, L6 and L7, with those at L1 and L6 extending into β -sheets B1 and B4. The mutation L1540P appears to disrupt the sheet structure of the domain, effecting a large shift in loop L7 and helix A2. Only G1505R and L1540P exhibit pronounced thermal fluctuations, and both showed the largest overall deviations from the model, further

highlighting a loss of structural stability. While the change for S1506L appears less dramatic, it involves large displacements of loops L3, L5 and L7, and unravelling at the end of helix A3. We speculate that the phenotype characteristic of group I mutations results from structural instability of the A2 domain, as best exemplified by G1505R and L1540P.

There are limitations that must be kept in mind when assessing the results of our simulations. We have used a homology model as no crystal structure is available. Significant errors in the model would affect the outcome of simulations. Conformational changes that occur on longer time scales would not be observed during our 10-ns simulations. This is especially true for mutations affecting core regions of the protein, such as L1540P, I1628T and L1503Q. The simulations started from a well-folded state. It is possible that this is never achieved in the cell, and that the actual structure is significantly misfolded. Also, we have simulated the isolated A2 domain, not the multimeric form of VWF. Fortunately, the residues at which the domain terminates are far from the crevice leading to the proteolysis site. For group II mutations which involve excess proteolysis, it is necessary to include the protease itself to fully understand the molecular basis of the phenotype. Notwithstanding these limitations, it was gratifying that many structural changes could be rationalized by way of altered interactions introduced by the mutations.

Supplementary material

The coordinates of the WT model in PDB format are available as supplementary material.

Acknowledgements We thank Z. Jia and the Department of Biochemistry, and A. Becke for providing access to computing resources. JJS acknowledges support from a Canadian Institute of Health Research doctoral research award. LAO acknowledges support from a Canadian Blood Services graduate fellowship. DL is the recipient of a Canada Research Chair in Molecular Hemostasis and a Career Investigator Award from the Heart and Stroke Foundation of Ontario. DFW acknowledges support from the Canada Research Chairs program. This study has been funded in part by grants from the Canadian Institutes of Health Research (MOP42467) and the Heart and Stroke Foundation of Ontario (T4421).

References

- Sadler JE (1991) *J Biol Chem* 266:22777–22780
- Ruggeri ZM, Ware J (1992) *Thromb Haemost* 67:594–599
- Berliner S, Niiya K, Roberts JR, Houghten RA, Ruggeri ZM (1988) *J Biol Chem* 263:7500–7505
- Voorberg J, Fontijn R, Calafat J, Janssen H, van Mourik JA, Pannekoek H (1991) *J Cell Biol* 113:195–205
- Marti T, Rosselet SJ, Titani K, Walsh KA (1987) *Biochemistry* 26:8099–8109
- Hoyer LW, Shainoff JR (1980) *Blood* 55:1056–1059
- Federici AB, Bader R, Pagani S, Colibretti ML, De Marco L, Mannucci PM (1989) *Br J Haematol* 73:93–99
- Schneppenheim R, Budde U, Oyen F, Angerhaus D, Aumann V, Drewke E, Hassenpflug W, Haberle J, Kentouche K, Kohne E, Kurnik K, Mueller-Wiefel D, Obser T, Santer R, Sykora K-W (2003) *Blood* 101:1845–1850
- Kokame K, Matsumoto M, Soejima K, Yagi H, Ishizashi H, Funato M, Tamai H, Konno M, Kamide K, Kawano Y, Miyata T, Fujimura Y (2002) *Proc Natl Acad Sci* 99:11902–11907
- Dent JA, Berkowitz SD, Ware J, Kasper CK, Ruggeri ZM (1990) *Proc Natl Acad Sci* 87:6306–6310
- Levy GG, Nichols WC, Lian EC, Foroud T, McClintick JN, McGee BM, Yang AY, Siemieniak DR, Stark KR, Gruppo R, Sarode R, Shurin SB, Chandrasekaran V, Stabler SP, Sabio H, Bouhassira EE, Upshaw Jr JD, Ginsburg D, Tsai HM (2001) *Nature* 413:488–494
- Verweij CL, Quadt R, Briet E, Dubbeldam K, van Ommen GB, Pannekoek H (1988) *J Clin Invest* 81:1116–1121
- Lyons SE, Cooney KA, Bockenstedt P, Ginsburg D (1994) *Blood* 83:1551–1557
- Sadler JE (1994) *Thromb Haemost* 71:520–525
- Lyons SE, Bruck ME, Bowie EJ, Ginsburg D (1992) *J Biol Chem* 267:4424–4430
- Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) *Annu Rev Biophys Biomol Struct* 29:291–325
- Jenkins PV, Pasi KJ, Perkins SJ (1998) *Blood* 91:2032–2044
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) *Nucleic Acids Res* 25:3389–3402
- Shindyalov IN, Bourne PE (1998) *Protein Eng* 11:739–747
- Smith TF, Waterman MS (1981) *J Mol Biol* 147:195–197
- Kelley LA, MacCallum RM, Sternberg MJ (2000) *J Mol Biol* 299:499–520
- Fischer D (2000) *World Sci* 119–130
- Jones DT (1999) *J Mol Biol* 287:797–815
- Needleman SB, Wunsch CD (1970) *J Mol Biol* 48:443–453
- MODELLER, version 6.0 (2001)<http://www.salilab.org/modeller/modeller.html>
- Rychlewski L, Jaroszewski L, Li W, Godzik A (2000) *Protein Sci* 9:232–241
- Bates P, Jackson RM, Sternberg MJ (1997) *Proteins Suppl* 1:59–67
- Bates P, Kelley LA, MacCallum RM, Sternberg MJ (2001) *Proteins Suppl* 5:39–46
- Karplus K, Barrett C, Hughey R (1998) *Bioinformatics* 14:846–856
- Eddy SR (2001)<http://hmmer.wustl.edu/>
- Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer ELL (2002) *Nucleic Acids Res* 30:276–280
- Grubmueller H (1996)<http://www.mpibpc.gwdg.de/abteilungen/071/solvate/docu.html>
- Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) *J Comput Phys* 151:283–312
- MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher III WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) *J Phys Chem B* 102:3586–3616
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) *J Chem Phys* 79:926–935
- Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) *J Comput Phys* 23:327–341
- Darden T, York D, Pedersen L (1993) *J Chem Phys* 98:10089–10092
- Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) *J Chem Phys* 103:4613–4621
- Tuckerman M, Berne BJ, Martyna GJ (1992) *J Chem Phys* 97:1990–2001
- Lee B, Richards FM (1971) *J Mol Biol* 55:379–389
- Kabsch W, Sander C (1983) *Biopolymers* 22:2577–2637

42. Lee JO, Bankston L, Arnaut MA, Liddington R (1995) *Structure* 15:1333–1340
43. Sauder JM, Arthur JW, Dunbrack RL (2000) *Proteins* 40:6–22
44. Jones DT (1999) *J Mol Biol* 292:195–202
45. Rost B, Sander C (1993) *J Mol Biol* 232:584–599
46. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ (1998) *Bioinformatics* 14:892–893
47. Sippl MJ (1993) *Proteins* 17:355–362
48. Honig B, Nicholls A (1995) *Science* 268:1144–1149
49. Tramontano A, Leplae R, Morea V (2001) *Proteins Suppl* 5:22–38
50. Edwards YJ, Perkins SJ (1995) *FEBS Lett* 358:283–286
51. Edwards YJ, Perkins SJ (1996) *J Mol Biol* 260:277–285
52. Lehninger AL, Nelson DL, Cox MM (1993) *Principles of biochemistry*. Worth Publishers, New York
53. Lankhof H, Damas C, Schiphorst ME, Ijsseldijk MJW, Bracke M, Furlan M, de Groot PG, Sixma JJ, Vink T (1999) *Thromb Haemost* 81:976–983
54. Kokame K, Matsumoto M, Fujimura Y, Miyata T (2004) *Blood* 103:607–612