Supporting Information

Shear-Induced Unfolding Activates von Willebrand Factor A2 Domain for Proteolysis

Carsten Baldauf*,†, Reinhard Schneppenheim‡, Wolfram Stacklies*,†, Tobias Obser†, Antje Pieconka§, Sonja Schneppenheim§, Ulrich Budde¶, Jing Zhou*,†, and Frauke Gräter*,†,#

* CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, P.R. China
† EML Research, Heidelberg, Germany
‡ Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
§ Coagulation Lab, AescuLabor Hamburg, Hamburg, Germany
# Max-Planck-Institute for Metals Research, Stuttgart, Germany

Corresponding authors:
Carsten Baldauf
CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, 200031 Shanghai, P.R. China
Tel: +86 2154920475; Fax: +86 2154920451
E-mail: carsten@picb.ac.cn

Reinhard Schneppenheim
Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Martinistraße 52, D-20246 Hamburg, Germany
Tel: +49 40428034270; Fax: +49 40428034601
E-mail: schneppenheim@uke.de

Frauke Gräter
EML Research gGmbH - Villa Bosch, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany
Tel: +49 6221533267; Fax: +49 6221533298
E-mail: frauke.graeter@eml-r.villa-bosch.de

Carsten Baldauf and Reinhard Schneppenheim contributed equally to this work.
**Section A: MD simulation setup**

All standard force probe MD simulations and part of the analysis were carried out with the Gromacs suite of programs (version 3.3.1) [1, 2]. The OPLS all atom force field was used for the protein [3, 4]. The proteins were solvated in dodecahedral boxes with at least 7,500 TIP4p water molecules [5], and periodic boundary conditions were applied. The typical protonation states at pH 7 were chosen for ionizable groups of the peptide. The necessary amount of counter-ions (Cl⁻ and Na⁺) was added to ensure a neutral system. Prior to free MD simulations, steepest descent energy minimizations and position restrained MD simulations with heavy atom positions restrained with a harmonic potential using a spring constant of 1000 kJ/mol·nm² (100 ps) were performed. Temperature (300 K) and pressure (1 bar) were coupled to a Nosé-Hoover thermostat [6, 7] and a Parrinello-Rahman barostat [8, 9], using time constants of 0.1 ps and 1 ps, respectively. Non-bonded interactions were considered within a cut-off of 1 nm, and long-range electrostatic interactions were calculated using the Particle-Mesh-Ewald algorithm [10, 11]. Constraints were applied by the LINCS algorithm [12]. A time step of 2 fs was used for integration. The wild type and mutant A2 models were simulated three times each for 30 ns and with different seeds for the initial velocity generation.

Three independent force-probe MD simulations were performed on a truncated VWF A2 model (residues 1492 to 1670), each ~26 ns in length. Harmonic springs with spring constants of 500 kJ/(mol nm²) were moved away from each other with a velocity of 1.25 nm/ns. To restrict the system size along the pulling direction, after partial unfolding the residues 1636 to 1670 of A2 were removed, water was added to the system, and the force-probe MD simulations were continued.

Pymol (http://www.pymol.org) [13], VMD (http://www.ks.uiuc.edu/Research/vmd/) [14] and POV-Ray (http://www.povray.org) were used for visualization.
Section B: Force distribution analysis (FDA)

For FDA two snapshots of the unfolding trajectory were used. As a folded state we used an intermediate (cf. Fig. 2, snapshot 2) with an unfolded terminal helix $\alpha 6$. The partially unfolded state (cf. Fig. 2, snapshot 4) models the cleavage ready VWF A2 domain. Constant force of 10 and 100 pN, respectively, for the relaxed and stretched state, was applied in opposing direction to both termini. Each of the two systems was equilibrated under the respective constant force for 30 ns. For both systems, the all atom RMSD to the starting structure remained below 0.35 nm for both pulling forces, indicating that the system is able to bear the mechanical stress within this time scale without rupture (Supplementary Fig. S3). In the following, 10 simulations for the folded and 20 simulation for the unfolded state were performed for 30 ns (folded) and 15 ns (unfolded) each, starting with different random velocities.

We used the FDA code [15] for Gromacs 4.0 [16] to write out forces $F_{ij}$ between each atom pair $i$ and $j$. Forces were averaged over the total simulation time of 00 ns per system, respectively, sufficient to obtain converged averages. Changes in forces, $\Delta F$, were then obtained as the difference in pair wise forces between the systems pulled with 10 and 100 pN. Residue wise forces $F_{uv}^{res}$ were obtained by summing up forces $F_{ij}$ for all pairs of atoms $i$ and $j$ in residues $u$ and $v$, where atom $i$ and atom $j$ must not be part of the same residue. The absolute sum $\Delta F_{uv}^{res} = \sum_v |\Delta F_{uv}^{res}|$ reflects the changes in strain acting on a single residue and was used to color code force distribution onto the protein backbone. Strain along the backbone was measured as the sum of all bonded interactions (bonds + angles + dihedrals) between adjacent residue pairs. Our MD simulations for the folded state use LINCS [12] to constrain bond length, and thus no forces for bonds could be calculated for this state. No constraints were used for the unfolded state.

Fig. S1. (A) Superposition of the average structures under 10 and 100 pN in the folded state used for FDA. Structures are averages over 300 ns, respectively. (B) Superposition of the average structures under 10 and 100 pN of the unfolding intermediate. Structures are averages over 300 ns, respectively.
Section C: Homology modeling

The sequences of the VWF A domains have a residue identity of 20 to 25%. Based on multiple sequence alignments and structural alignments we created a homology model of the VWF A2 domain (residues 1488 to 1676 of human VWF) and the mutant A2 domain (N1493C and C1670S) from a human VWF A1 X-ray structure (PDB: 1AUQ).

The search for similar sequences was performed in two steps: a fast search with a generalized Fasta methodology and an evaluation based on E-values and Z-scores [17]. The structural model comprised residues 1488 to 1676 of human VWF, and therefore a longer sequence than the one used by [18]. Finally, pdb-structures (1AUQ, 1ATZ, 1IJB, 1U0O, 2ADF, 1SHU, 1PT6) were selected and subjected to a structural alignment (Fig. S1). The resulting sequence alignment is shown in the Supplementary Information. 20 homology models were created from 20 randomized starting configurations, based on the VWF A1 domain (1AUQ). Structures were evaluated on the basis of energies from the Amber99 force field as implemented in MOE. Homology modeling was performed using the molecular operation environment MOE (2007.9, Chemical Computing Group CCG).

Based on the model of the A2 domain, the A2 double mutant N1493C/C1670S was generated. A disulfide bridge was introduced between the termini by the N1493C mutation enabling a link between C1493 and C1669. To maintain a constant content of cysteine residues, known to be beneficial for protein expression, a second mutation C1670S was introduced. The models were validated by molecular dynamics (MD) simulations.

The homology models of wild type and mutant VWF A2 domain were characterized with ProSA 2003 [19] (Fig. S2A) and equilibrium Molecular Dynamics (MD) simulations. Within the 30 ns simulation time for each of the three independent trajectories the structures converged fast to a backbone root mean square deviation (rmsd) between 0.2 and 0.25 nm (Fig. S2B and C). The agreement with the previous model and the overall high stability indicate the quality of this A2 model and its appropriateness for the subsequent studies. The coordinates of the models are available in PDB format as Supporting Information or upon request from the authors.
| A2  | LGPKRN...SMVL...DFVE...DQI...EG1...E2N...SF...KFE...F1...E1...GQDS...1HVTV...LOY |
| 1AUQ | DISEP...HDFVIC...SRL...LFL...LDG...S...S...EAE...F1...LAF...V...D...M...ER...L...R...SQKWV...AVVY |
| 1ATZ | DCSQP...L...L...G...S...S...S...F...P...S...F...E...D...M...K...S...F...A...K...I...G...P...R...T...O...V...S...LOY |
| 1UB | SEPLHDFVIC...S...R...L...FL...L...D...G...S...S...E...A...E...F1...L...A...F...V...D...M...E...R...L...R...S...Q...K...W...V...A...V...Y |
| 1UO | F...C...K...L...L...D...G...S...S...E...A...E...F1...L...A...F...V...D...M...E...R...L...R...S...Q...K...W...V...A...V...Y |
| 2ADF | DCSQP...L...L...G...S...S...S...F...P...S...F...E...D...M...K...S...F...A...K...I...G...P...R...T...O...V...S...LOY |
| 1SHU | SCRAFDLYFVIDKGS...V...N...W...IE...I...N...F...Y...Q...Q...L...A...ER...F...V...S...P...E...M...R...L...S...F...I...V |
| 1PT6 | QLD...I...V...L...D...G...S...N...I...Y...P...W...D...S...V...T...A...L...D...N...L...L...K...M...D...I...G...P...K...Q...T...V...G...I...Q|

### Fig. S2. Multiple sequence alignment used as basis for homology modeling.


Fig. S3. (A) Verification of the Homology Model with ProSA 2003: The energy analysis is smoothed with a window size of 30 aa. Characterizing the model with ProSA-Web shows a Z-score for the raw model of -6.99, and of -8.11 for the model after 10 ns MD simulation. The Z-score for the structure model published by Sutherland et al. is -7.84. (B) Backbone rmsd of the wild type A2 domain monitored in three independent 30 ns MD simulations. (C) Backbone rmsd of the mutant A2 (N1493C/C1670S) domain monitored in three independent 30 ns MD simulations.

Secondary structure organization of the VWF A2 domain model:

\[ \begin{align*}
\beta_1 & (L1497 \text{ to } E1504), \alpha_1 (E1511 \text{ to } Q1526), \beta_2 (I1535 \text{ to } Y1542), \beta_3 (V1546 \text{ to } P1551), \alpha_2 (D1560 \text{ to } R1566), \alpha_3 (T1578 \text{ to } D1587), \beta_4 (P1601 \text{ to } T1608), \alpha_4\text{-less loop (R1618 to G1621), } \\
\beta_5 (Q1624 \text{ to } V1630), \alpha_5 (Q1635 \text{ to } R1641), \beta_6 (P1648 \text{ to } I1651), \alpha_6 (F1654 \text{ to C1670}).
\end{align*} \]

Fig. S4. Superposition of 3GXB (silver) and the homology model (orange), the main chain RMSD is 0.189 nm. Selected structural features are shown (grey carbons: 3GXB, green carbons: homology model): vSS: the C terminal vicinal disulfide bond is not predicted in the model; \( \beta \text{VIIa} \) turn with a \( \text{cis} \) peptide bond in the A2 X-ray structure (\( \beta \text{VIIa} \) backbone torsion angles: \( \phi_1=-41.6; \psi_1=129.8; \omega=3.2; \phi_2=-79.7; \psi_2=1.7 \)) and a \( \beta \text{III} \) (\( \beta \text{III} \) \( \phi_1=-70; \psi_1=-50; \omega=180; \phi_2=-60; \psi_2=-50, \text{ ref. [20]} \)) turn featuring a \( \text{trans} \) peptide bond in the homology model; CS: the position and conformation of the cleavage site is well predicted.
Section D: Force induced unfolding of the A2 domain model

Fig. S5. The force profiles for three independent force-probe MD simulations of our VWF A2 domain model are shown. After extending the protein chain to 15 nm, the simulations were continued with the unfolded C terminal part (sequence numbers 1636 and higher) being cut off. Selected snapshots are shown as cartoon; the cleavage site is shown in green; the fully unfolded C-terminal fragments in 2, 3, and 4 are omitted for clarity.

Section E: Further provided supporting information

Dataset S1. Homology model of the A2 domain including VWF residues 1488 to 1676 in PDB-format.

Video S1. Visualization of a VWF A2 Domain Force Probe MD Simulation:
The N (blue sphere) and C terminus (red sphere) are pulled apart from each other, the secondary structure elements (red: helices; yellow: strands) are stepwise peeled off until the Tyr1605—Met1606 cleavage site (green) is uncovered. A part of the extended and unfolded C terminus is removed in order to save computing time. [xvid4 encoded, AVI container]
Section F: References for the supporting information


