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 dodecahedric 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·nm^2 (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 α6. 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, ΔF, 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 $\Delta F_u^{res} = \sum_v \left| \Delta F_{uv}^{res} \right|$ 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.

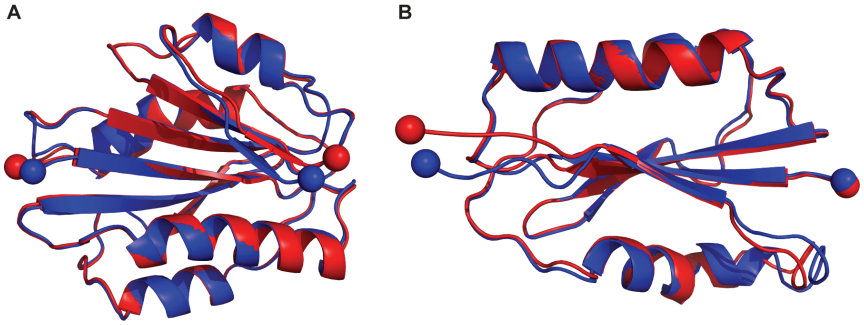

Fig. S1. (**A**) Superposition of the average structures under 10 and 100 pN in the folded state used for FDA. Structures are averages over 300 ns, respectively. (**B**) Superposition of the average structures under 10 and 100 pN of the unfolding intermediate. Structures are averages over 300 ns, respectively.

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.

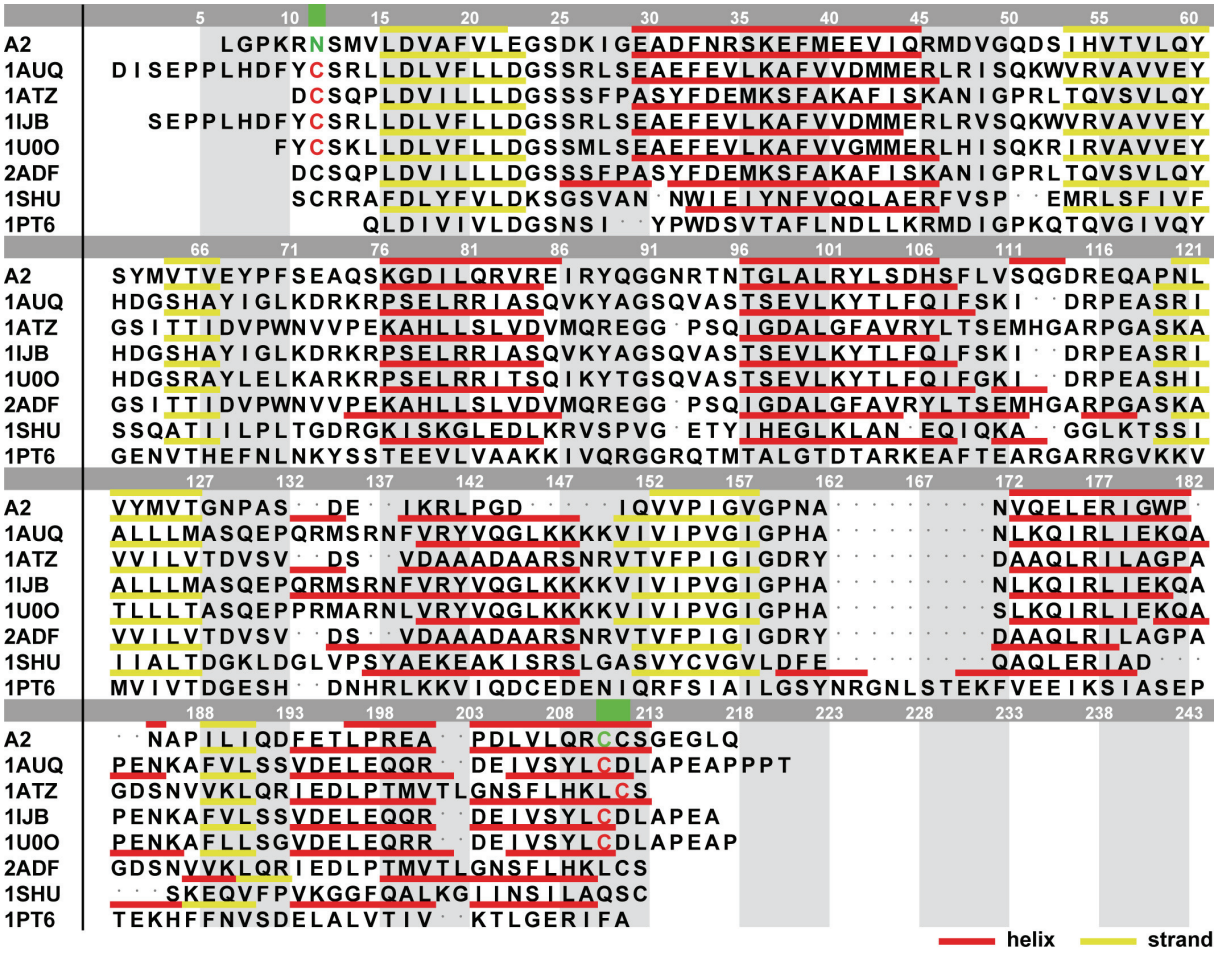

Fig. S2. Multiple sequence alignment used as basis for homology modeling. **1AUQ:** A1 domain of von Willebrand factor (Emsley J, Cruz M, Handin R, Liddington R (1998) J Biol Chem 273: 10396-10401). **1ATZ:** Human von Willebrand factor A3 domain (Huizinga EG, Martijn van der Plas, R, Kroon J, Sixma JJ, Gros P (1997) Structure 5: 1147-1156). **1IJB:** The von Willebrand factor mutant (I546V) A1 domain (Fukuda K, Doggett TA, Bankston LA, Cruz MA, Diacovo TG, Liddington RC (2002) Structure 10: 943-950). **1U0O:** The mouse von Willebrand Factor A1-botrocetin complex (Fukuda K, Doggett T, Laurenzi IJ, Liddington RC, Diacovo TG (2005) Nat Struct Mol Biol 12: 152-159). **2ADF:** Crystal Structure and Paratope Determination of 82D6A3, an Antithrombotic Antibody Directed Against the von Willebrand factor A3-Domain (Staelens S, Hadders MA, Vauterin S, Platteau C, De Maeyer M, van Hoorelbeke K, Huizinga EG, Deckmyn H (2006) J Biol Chem 281: 2225-2231). **2SHU:** Crystal Structure of the von Willebrand factor A domain of human capillary morphogenesis protein 2: an anthrax toxin receptor (Lacy DB, Wigelsworth DJ, Scobie HM, Young JAT, Collier RJ (2004) Proc Natl Acad Sci USA 101: 6367-6372). **1PT6:** I domain from human integrin alpha1-beta1 (Nymalm Y, Puranen JS, Nyholm TKM, Kapyla J, Kidron H, Airenne TT, Heino J, Slotte JP, Johnson MS, Salminen TA (2004) J Biol Chem 279: 7962-7970).

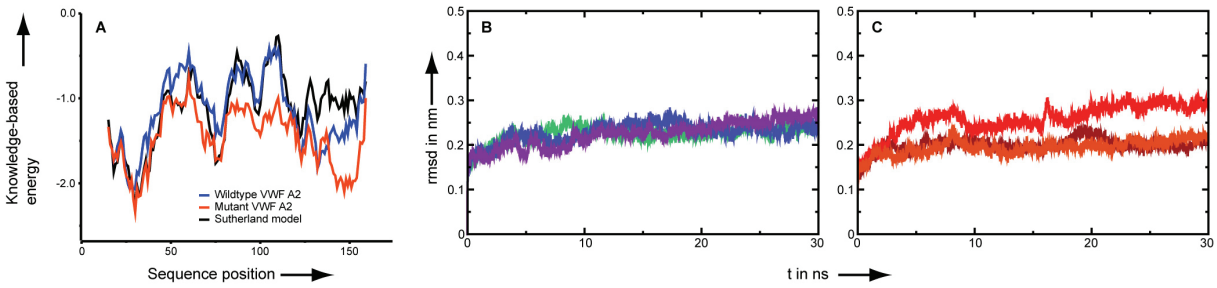

Fig. S3. (**A**) Verification of the Homology Model with ProSA 2003: The energy analysis is smoothed with a window size of 30 aa. Characterizing the model with ProSA-Web shows a Z-score for the raw model of -6.99, and of -8.11 for the model after 10 ns MD simulation. The Z-score for the structure model 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:

β1 (L1497 to E1504), α1 (E1511 to Q1526), β2 (I1535 to Y1542), β3 (V1546 to P1551), α2 (D1560 to R1566), α3 (T1578 to D1587), β4 (P1601 to T1608), α4-less loop (R1618 to G1621), β5 (Q1624 to V1630), α5 (Q1635 to R1641), β6 (P1648 to I1651), α6 (F1654 to C1670).

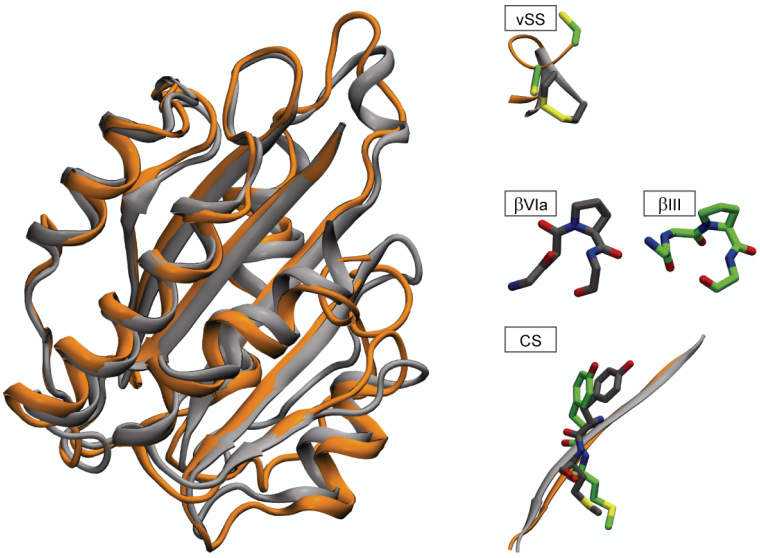

Fig. S4. Superposition of 3GXB (silver) and the homology model (orange), the main chain RMSD is 0.189 nm. Selected structural features are shown (grey carbons: 3GXB, green carbons: homology model): **vSS**: the C terminal vicinal disulfide bond is not predicted in the model; **βVIa** turn with a *cis* peptide bond in the A2 X-ray structure (βVIa backbone torsion angles: φ_1=-41.6; ψ_1=129.8; ω=3.2; φ_2=-79.7; ψ_2=1.7) and a **βIII** (βIII φ_1=-70; ψ_1=-50; ω=180; φ_2=-60; ψ_2=-50, ref. [20]) turn featuring a *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

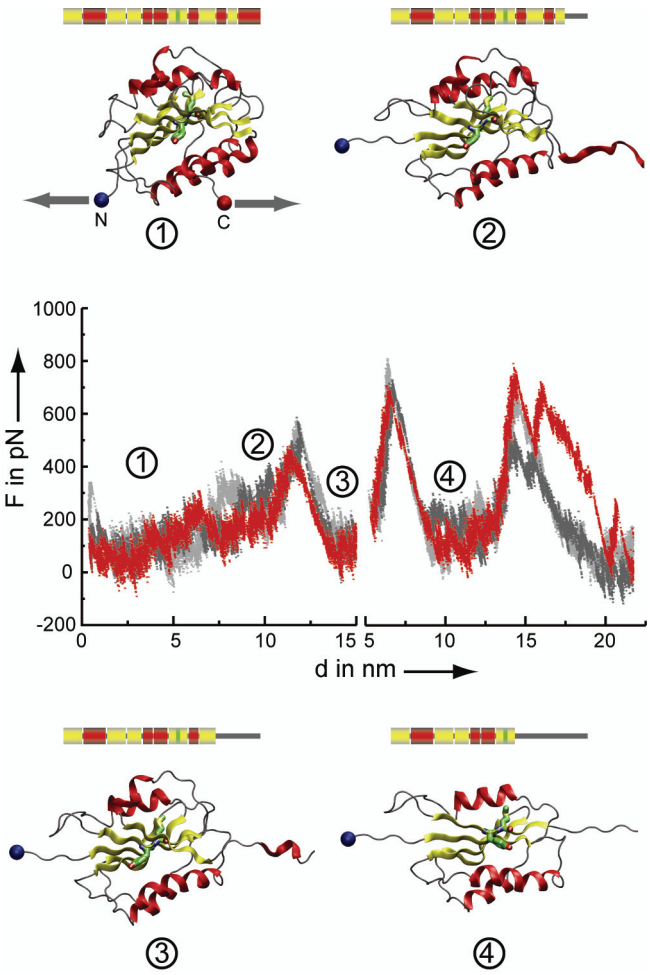

Fig. S5. The force profiles for three independent force-probe MD simulations of our VWF A2 domain model are shown. After extending the protein chain to 15 nm, the simulations were continued with the unfolded C terminal part (sequence numbers 1636 and higher) being cut off. Selected snapshots are shown 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 of 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

1 Berendsen HJC, van der Spoel D, van Drunen R. Gromacs - A message-passing parallel molecular-dynamics implementation. *Comput Phys Commun*. 1995; **91**: 43-56.
2 Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J Mol Model*. 2001; **7**: 306-17.
3 Jorgensen WL, Ulmschneider JP, Tirado-Rives J. Free energies of hydration from a generalized Born model and an ALL-atom force field. *J Phys Chem B*. 2004; **108**: 16264-70.
4 Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparameterization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B*. 2001; **105**: 6474-87.
5 Lawrence CP, Skinner JL. Flexible TIP4P model for molecular dynamics simulation of liquid water. *Chem Phys Lett*. 2003; **372**: 842-7.
6 Hoover WG. Canonical dynamics - Equilibrium phase-space distributions. *Phys Rev A*. 1985; **31**: 1695-7.
7 Nosé S. A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys*. 1984; **52**: 255-68.
8 Nosé S, Klein ML. Constant pressure molecular dynamics for molecular systems. *Mol Phys*. 1983; **50**: 1055-76.
9 Parrinello M, Rahman A. Polymorphic transitions in single crystals - A new molecular dynamics method. *J Appl Phys*. 1981; **52**: 7182-90.
10 Darden T, York D, Pedersen L. Particle mesh Ewald - An Nlog(N) method for Ewald sums in large systems. *J Chem Phys*. 1993; **98**: 10089-92.
11 Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995; **103**: 8577-93.
12 Hess B, Bekker H, Berendsen HJC, Fraaije J. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*. 1997; **18**: 1463-72.
13 DeLano WL. The PyMOL Molecular Graphics System. Palo Alto, CA, USA: DeLano Scientific LLC, 2008.
14 Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996; **14**: 33-8.
15 Stacklies W, Vega MC, Wilmanns M, Gräter F. Mechanical Network in Titin Immunoglobulin from Force Distribution Analysis. *PLoS Comput Biol*. 2009; **5**: e1000306.
16 Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 2008; **4**: 435-47.
17 Pearson WR. Effective protein sequence comparison. In: Doolittle RF, ed. *Methods Enzymol*. San Diego: Academic Press, 1996, 227-58.
18 Sutherland JJ, O'Brien LA, Lillicrap D, Weaver DF. Molecular modeling of the von Willebrand factor A2 domain and the effects of associated type 2A von Willebrand disease mutations. *J Mol Model*. 2004; **10**: 259-70.
19 Sippl MJ. Recognition of errors in 3-dimensional structures of proteins. *Proteins*. 1993; **17**: 355-62.
20 Möhle K, Gussmann M, Hofmann H-J. Structural and energetic relations between beta turns. *J Comput Chem*. 1997; **18**: 1415-30.