

## Supplementary Information

for the article

### **The Membrane Anchor of the Transcriptional Activator SREBP is Characterized by Intrinsic Conformational Flexibility**

by Rasmus Linser, Nicola Salvi, Rodolfo Briones, Petra Rovó, Bert L. de Groot, and Gerhard Wagner

#### **Experimental:**

##### *Protein expression*

Expression of the SREBP membrane anchor was achieved using cell-free expression of a pIVEX4d vector encoding amino acids M482 to V566 (full-length anchor) or to L530 (anchor as processed by S1P). Constructs without and with an N-terminal His<sub>8</sub> tag with factor X, TEV, and Thrombin cleavage sites were generated. Cell-free reactions were set up in accordance to the protocol by Schwarz et al. (1) using an A19 *E. coli* S30 cell extract and deuterated algal amino acid mix obtained from Sigma Aldrich or Cambridge Isotopes in the case of triply- (<sup>2</sup>D, <sup>13</sup>C, <sup>15</sup>N) and doubly (<sup>2</sup>D, <sup>15</sup>N) labeled amino acids, respectively, recombinantly expressed T7 RNA polymerase and pyruvate kinase purchased from Sigma Aldrich. We used single amino acids Gln, Trp, Cys, and Asn in an either doubly (<sup>13</sup>C, <sup>15</sup>N) labeled or unlabeled form. One sample employed combinatorial labeling (Cys, Asn, Leu doubly <sup>13</sup>C, <sup>15</sup>N labeled in combination with a doubly (<sup>2</sup>D, <sup>15</sup>N) labeled algal amino acid mix. Batch expression yielded on the order of 10 mg protein per 3 ml reaction mixture (using 3 ml slide-A-lyzers in a Petri dish containing 30 ml feeding mixture).

Cell-free expression pellets were centrifuged (30 min, 14 krpm) and washed twice by resuspension in 20 mM Tris pH 7.5, 100 mM NaCl, 5mM DTT and 1/100 protease inhibitor (buffer A).

Spontaneous refolding of the peptide in lyso myristoyl glycerophosphoglycerol (LMPG) micelles was achieved by infinite dilution (1:20 volume ratio) from a guanidium solution into 5% LMPG in buffer A, using 5-fold LMPG excess with respect to protein (m/m). Typically, 1 ml of 20 mg protein-containing buffer was refolded by dropwise dilution into 20 ml buffer A containing 100 mg LMPG. After buffer exchange with 0.08 % LMPG, pH 8.5, 5 mM β-ME, 20 mM Tris, 100 mM NaCl and 1/100 protease inhibitor, by three times concentration to 3.5 ml and redilution to 15 ml using Amicon centricons with a cutoff of 10 kDa, Talon resin (8 ml per 20 mg protein) was loaded by very slow passage of the solution, extensive wash with the same buffer and elution with the same buffer in the presence of 500 mM imidazole. Sample quality obtained upon reconstitution trials into nanodiscs or bicelles were insufficient for the requirements of subsequent NMR studies.

Even though the cell-free expressed protein could be purified to sufficient purity also without the use of histidine tags and metal-affinity chromatography, expression of

constructs obtained from deletion of the N-terminal histidine tag in pIVEX vectors resulted in approximately fivefold less yield compared with constructs bearing the optimized pIVEX4d N-terminal residues. Cleavage of the tag could be achieved when using a Thrombin cleavage site including a GSGS linker between this site and the anchor. This protease cleavage, however, was omitted for most preparations since, due to star activity of Thrombin, the benefits from spectral simplification did not outweigh the protein loss incurred, particularly after completion of resonance assignment.

The final NMR sample in a Shigemi tube was obtained by buffer exchange against PBS (pH 6.3, with 5 mM DTT and 1/100 protease inhibitor) and concentration to 280  $\mu$ l using spin concentrators of decreasing size.

S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methane-sulfonothioate (MTSL) spin-labeled single-cysteine mutants were produced using Ser to Cys mutants at positions S486 and S499 upon mutation of other Cys residues to Ala. Subsequent reduction of the spin label was pursued using 1 mM ascorbic acid.

### *Spectroscopy*

All NMR experiments were recorded at 37 °C on Bruker spectrometers of 800 or 600 MHz Larmor frequency and Avance III consoles equipped with Topspin 3.2 or a 750 MHz Bruker spectrometer with an Avance III console and Topspin 2.1. Additionally, a 500 MHz Varian spectrometer with a cryprobe and a 500 MHz Bruker spectrometer with a room temperature probe were used initially to support biochemistry and for screening of NMR conditions. For resonance assignments, HNCA, HNCACB, HNCO, and HNCOCA experiments and  $^{15}\text{N}$ -edited NOESY experiments were recorded on various deuterated,  $^{15}\text{N}$ , and  $^{13}\text{C}$ -labeled full-length and S1P-shortened (only HNCACB and HNCO) SREBP anchor constructs, using standard or non-uniform sampled acquisition.  $^{15}\text{N}$ -edited NOESY experiments were recorded using 90 ms NOESY mixing time on both constructs at 800 MHz.

Apart from most residues of N-terminal His<sub>8</sub> tags and their cleavage sites in uncleaved NMR samples, we could unambiguously assign 38 amino acids (85 % of all non-Pro residues) of the SP2 substrate, and 60 amino acids (78 % of all non-Pro residues) of the two-TM helix membrane anchor, see Figure 2A. The first five N-terminal residues, which include an S2P recognition site and are predicted to be flexible, non-membrane-embedded residues, could not be assigned, presumably due to fast H<sup>N</sup> exchange with solvent. The remainder of those residues unassigned proved impossible to identify due to fully protonated amino acids in the case of Gln, Trp, Cys, and Asn in combination with the long tumbling correlation times of the membrane protein constructs in a micellar environment. In the case of the S1P-unprocessed anchor, a large stretch within the second helix (Q541 – N553) could not be assigned unambiguously even upon partial use of combinatorial labeling, probably due to the degenerate chemical shifts of repetitive Leu or Val residues in addition to several Pro or non-deuterated (NCWQ) residues, which were generally not observable (QWLLPPVVLLN).

Dynamics experiments (hetNOE (steady-state NOE using 3 s  $^1\text{H}$  irradiation times),  $^{15}\text{N}$   $T_1$ , and  $^{15}\text{N}$   $T_2$ , on the shortened anchor only) were recorded at 800 MHz. All NMR data were processed using Topspin or NmrPipe (2) (for non-uniform sampled experiments) and further used in CCPNmr Analysis (3), Microsoft Excel, and Pymol (4). NOE data as

well as transverse and longitudinal relaxation rates were used to perform reduced spectral density mapping in Relax (5). Cross-correlated cross-relaxation experiments were performed using the methods of Pelupessy et al. (6, 7) The spectra were processed in NmrPipe, and peak intensities extracted using CCPNmr Analysis were used to calculate the cross-relaxation rates using self-made scripts in Python. General graphical illustration was pursued using Adobe Illustrator.

Preliminary structural models were obtained using chemical shifts (TALOS+) (8), NOE, and PRE data in combination with CNS (9).

#### *Molecular dynamics simulations (MD)*

MD simulations and analysis were performed with the Gromacs 5.0 (10) package. The starting peptide structure as obtained from CNS based on secondary-structure restraints and NOEs was inserted in equilibrated POPC, DLPC lipid bilayers using `g_membed` (11). It was also inserted in a pure water box (without bilayers). The solvent and the membranes were equilibrated around the peptide for 200 ns, keeping the peptide heavy atoms restrained, using a force constant of  $1000\text{-kJ}\cdot\text{mol}^{-1}\text{ nm}^2$ . The pressure was kept constant at 1 bar (Berendsen pressure coupling (12)) and the temperature was 37 °C (velocity rescale thermostat (13)). A Lennard-Jones cutoff of 1.0 nm was used and PME (14) for long-range electrostatics. The integration timestep was 2 fs. LINCS (15) and SETTLE (16) were used to constrain the bonds of molecules and water, respectively. The SPCE water model was used, together with the Amber99sb-ILDN (17) forcefield for the protein and the Cordomi (18) United Atoms forcefield for POPC lipids and the Stockholm lipids port for DLPC (19). For the production run, simulations of 1  $\mu\text{s}$  each were run for the peptide in POPC, DLPC and water. Root mean square fluctuations (RMSF) of the backbone atoms were calculated after fitting the trajectories to the backbone of amino acids 5 to 20 or 10 to 20. The secondary structure of the peptides in the simulations were estimated using DSSP (20). The hinge angle of the helical parts around the NP motif was calculated in each trajectory excluding the first 100 ns. The  $C^\alpha$  of residues 487, 488 and 498, 499 for the first helical part (before NP) and the  $C^\alpha$  of residues 502, 503 and 505, 506 for the last helical part (after the NP) were used to define the helix vectors. The relative extent of different types of secondary structure and RMSF over time were extracted from equidistant snapshots spaced by 0.4 ns using the Amber99sb-ILDN force field.

Circular Dichroism spectroscopy was pursued using a Jasco J-815 spectrometer.

#### *Bioinformatic analysis*

Helix packing prediction was pursued using `psypred` MEMPACK (21). In accordance with the hydrophobic helix surfaces lacking a zipper or other common packing sequences, helix packing prediction yielded no or multiple, poorly defined, and contradictive interaction surfaces. Only one (conserved) polar residue (Asn353) can be found in the second helix, whereas the first helix of SREBP1 contains two polar residues (Ser and Thr) near the cleavage site.

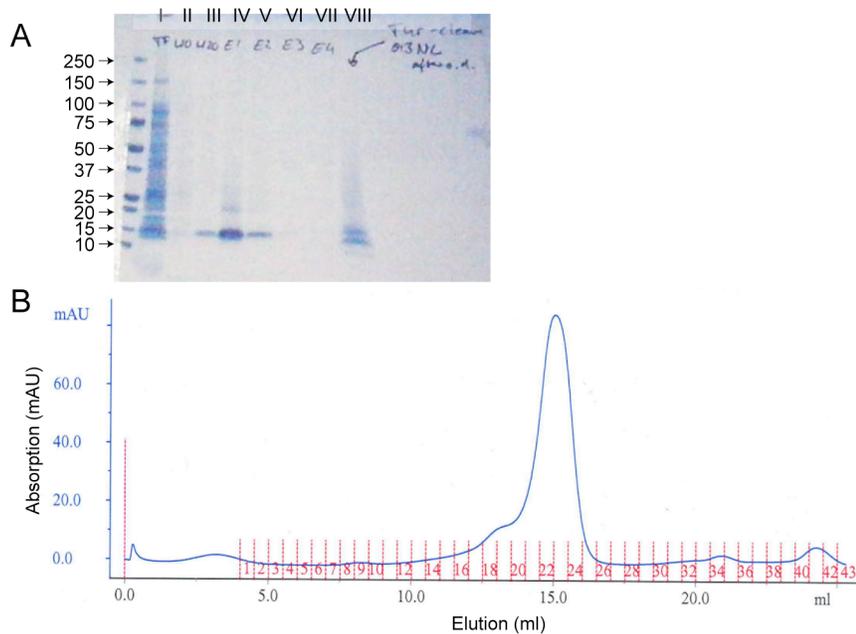
Even though a stable interaction between the two helices is unlikely also in the physiological case, at least temporary contacts are likely and cannot be excluded by the in-vitro data. Such an interaction, however, seems of low significance generally and for the analysis here, since the crucial step of S2P proteolysis occurs with the S1P processed

anchor. Consequently, we are here focusing on this S1P-cleaved S2P substrate in this work, which bears only the N-terminal helix.

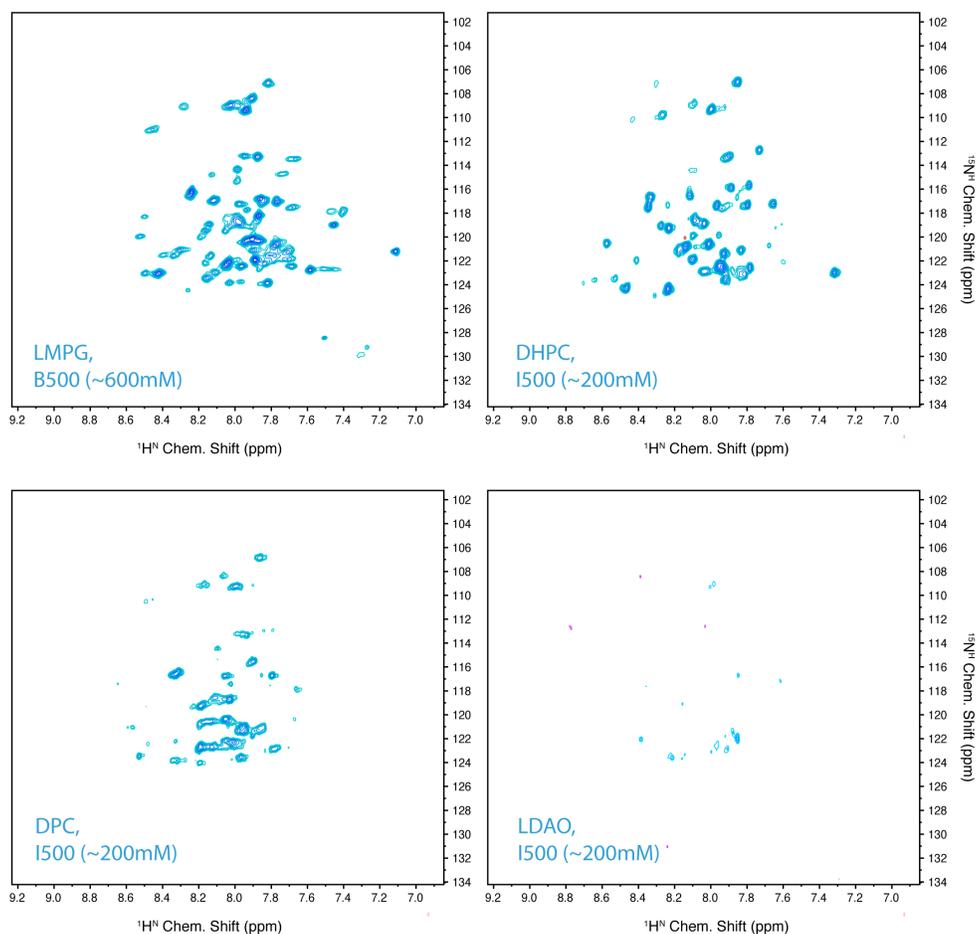
Primary sequences from arbitrarily selected single-TM proteins (22) for hydropathy plots were obtained from Uniprot. We selected non-precursor (non-RIP-substrate) proteins exclusively for the comparison: the transport protein comB (P36498), the disintegrin and metalloproteinase domain-containing protein 10 (ADAM10, Q10741), and the mast cell surface glycoprotein Gp49A (Q61450). Hydropathy characteristics (Kyte and Doolittle (23)) were determined using the EMBOSS Pepwindow framework (24). Topology was predicted using the JPRED prediction software (25).



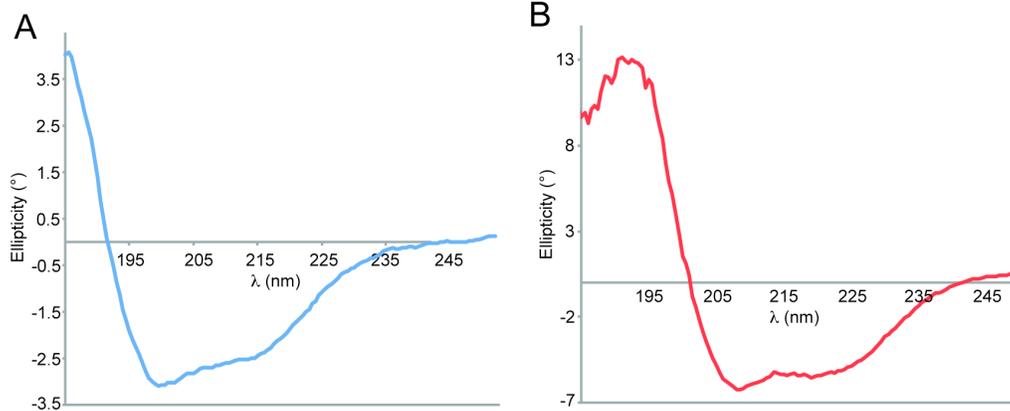
**Figure 1:** JPRED prediction of amino acid TM regions as in Figure 1 in the Main text, but using only the fragment shown as an input for the routine (brown: helix and less than 25% accessible to solvent, blue: random coil and more than 25% solvent accessible, yellow: coil and less than 25% accessible). No  $\beta$ -strand is predicted near the S1P cleavage site.



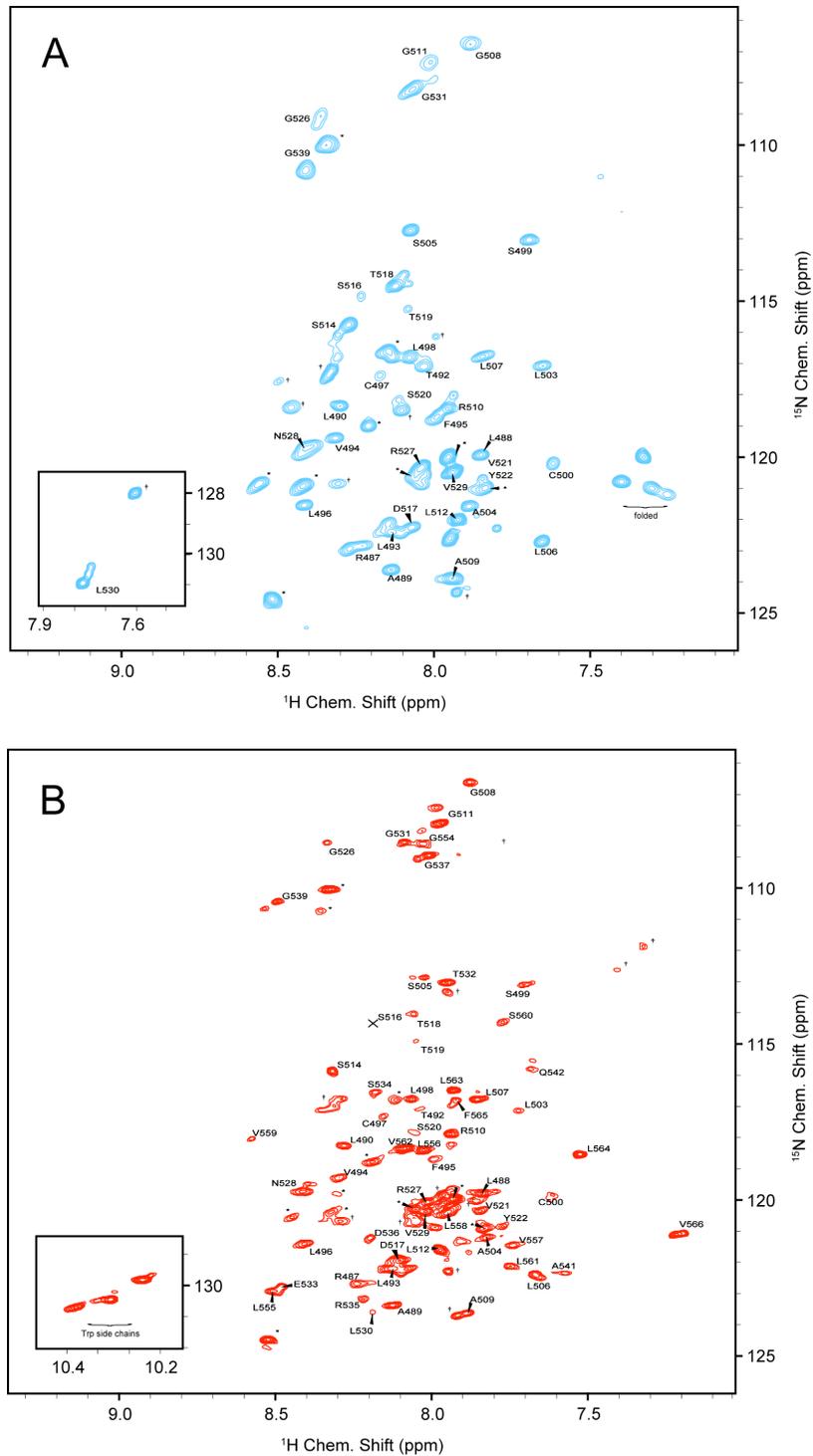
**Figure 2:** Typical outcome of SREBP purification. **A)** Metal affinity chromatography of the refolded cell-free SREBP full-length anchor precipitate. Lines I to VII represent throughflow, wash fractions with 0 and 20 mM imidazole, and 4 elution fractions of 1 CV each on gravity-column Nickel resin. Fraction VIII shows an (incomplete) Thrombin cleavage at room temperature. In later stages of the project, Talon resin was used rather than Nickel beads due to better compatibility with reducing agents. **B)** Monodisperse profile of the anchor in size exclusion chromatography on an analytical S200 column. The use of SEC was observed to provide hardly any sample improvement and was usually omitted in later stages of the project.



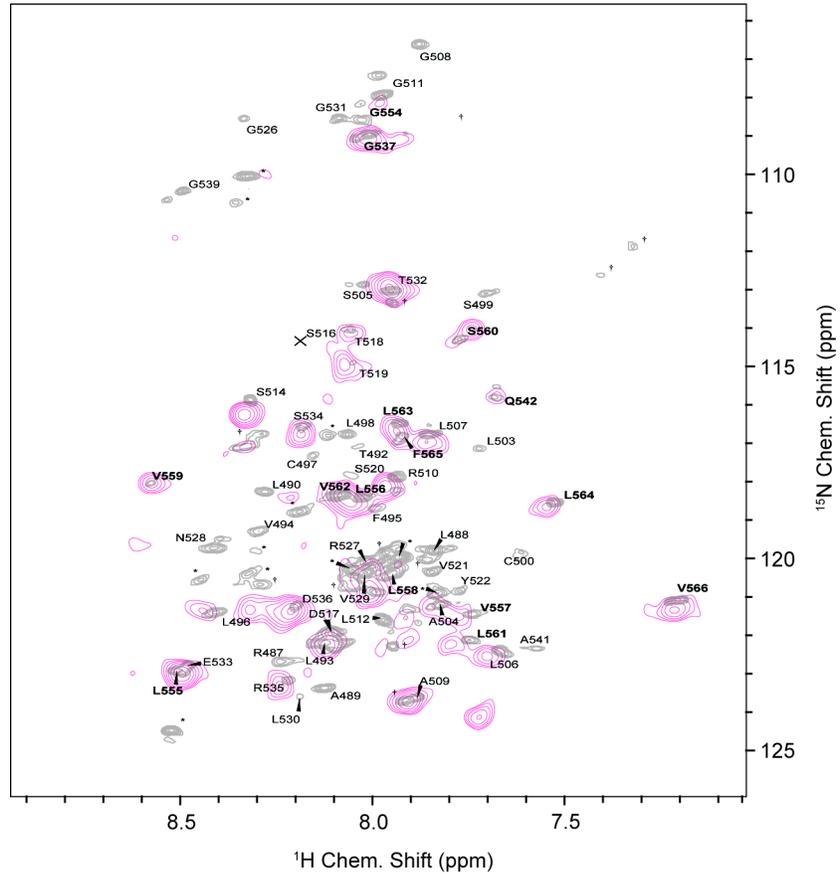
**Figure 3:** Detergent screens at 500 MHz proton Larmor frequency. Shown are spectra of the full-length SREBP anchor expressed without a histidine tag recorded in lyso myristoyl glycerophosphoglycerol (LMPG), dihexanoyl-glycerophosphocholin (DHPC), dodecyl phosphocholin, and lauryldimethylamine oxide (LDAO) in 7 h each. The spectrum in LMPG was recorded using a 600 mM sample and a room temperature probe, whereas the others were recorded on 200 mM samples using a spectrometer equipped with a cryoprobe.



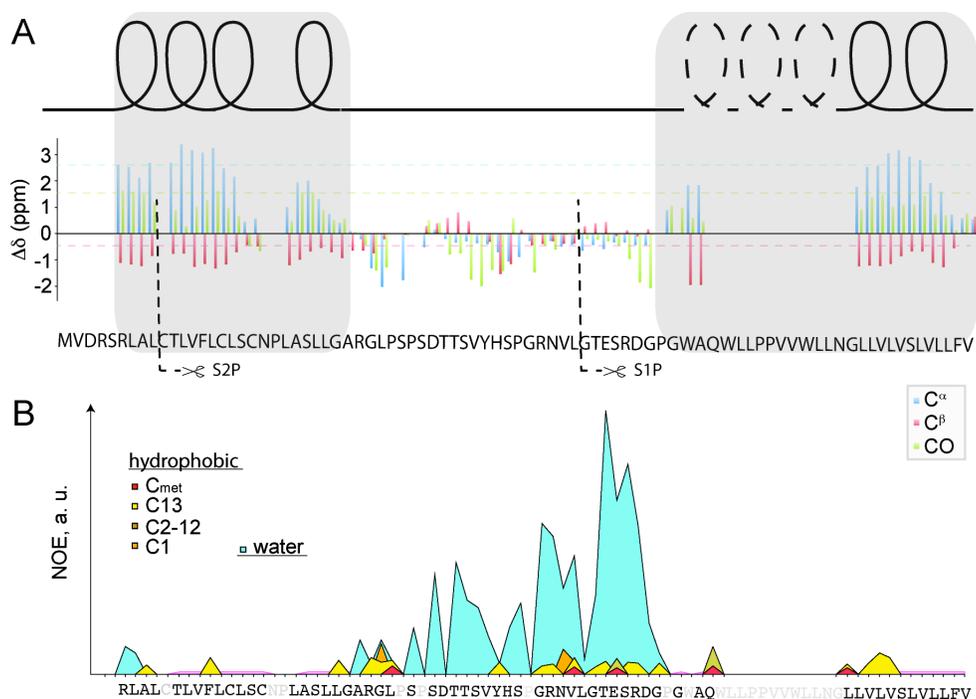
**Figure 4:** Circular Dichroism (CD) spectra of the SREBP membrane anchor. **A)** shows the anchor as shortened by S1P consisting of one TM helix and most of the linker. **B)** shows the anchor construct including both TM domains. Here, the  $\alpha$ -helical content is expectedly stronger than in A). LMPG concentrations amount to 0.03 % and 0.05 % for A) and B), respectively.



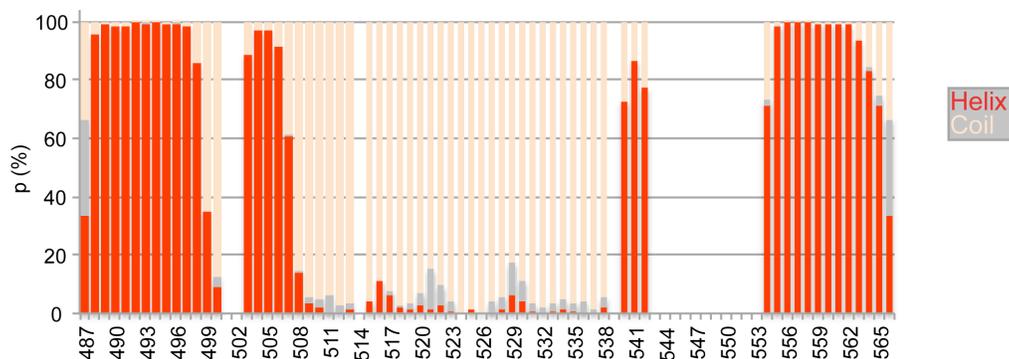
**Figure 5:** HSQC spectra with assigned residues at 800 MHz proton Larmor frequency. **A)** SREBP anchor as shortened by S1P, containing residues M482 to L530, **B)** complete anchor (M482 to V566). Asterisks mark residues assigned but belonging to the uncleaved N-terminal tag. Crosses mark unassigned residues. The S516 peak is below the contours shown in the plot.



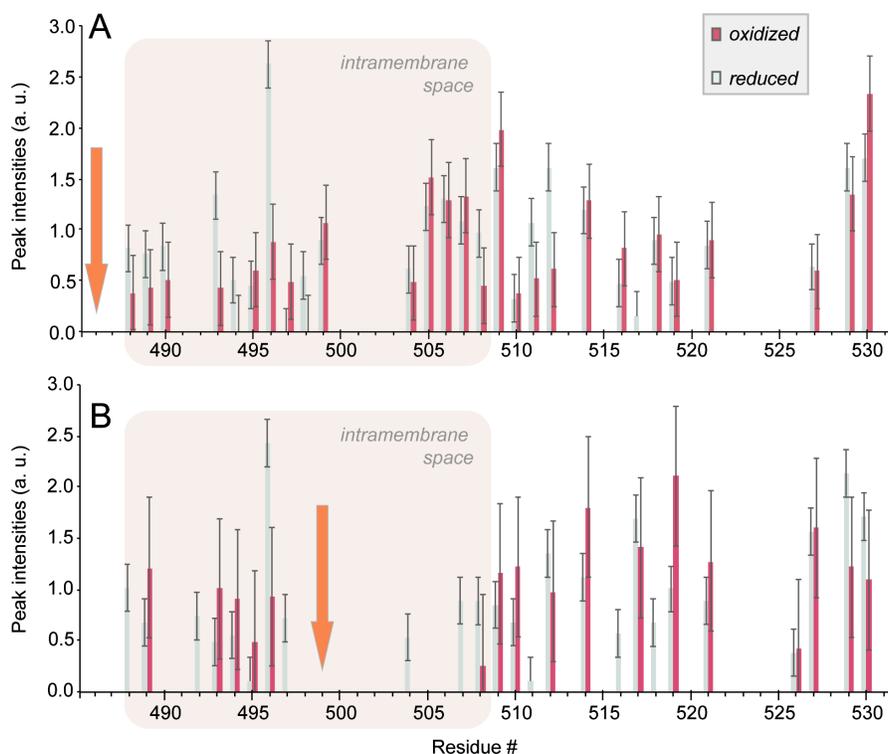
**Figure 6:** Overlay of spectra recorded on the full-length anchor (gray) and a single-cysteine mutant of the full-length anchor with an MTSL spin label just above the first transmembrane helix (magenta). The second-helix residues (bold) are not affected by paramagnetic attenuation, indicating the expected absence of packing interactions between helix 1 and 2. This behavior found in detergent micelles *in vitro* is expected in the physiological case in the course of transcriptional signaling, where the particularly long linker between the helices and S1P cleavage enable their independent traveling.



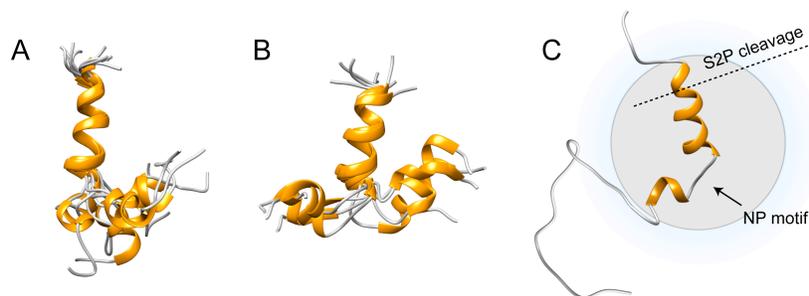
**Figure 7:** Secondary chemical shifts and water accessibility of the uncleaved SREBP anchor as in Main Manuscript Figures 2B and C. A long stretch in the second helix was unassignable due to sensitivity restrictions induced by the need for protonated C, W, Q, N amino acids upon otherwise deuterated cell-free expression.



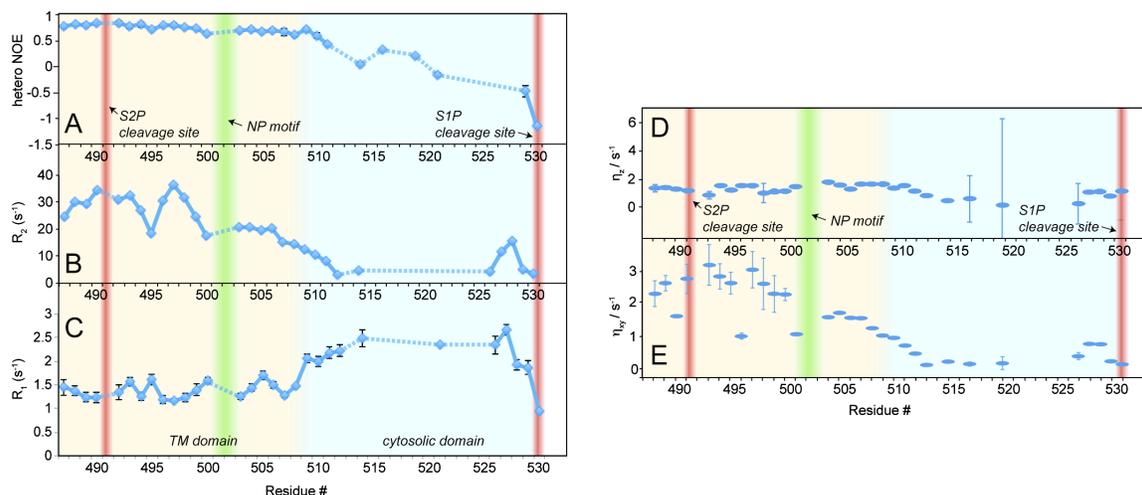
**Figure 8:** Secondary chemical structure as predicted by TALOS+. (8) The y-axis shows the prediction probability for different types of secondary structure based on  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}^{\text{H}}$ ,  $^{13}\text{C}^{\alpha}$ ,  $^{13}\text{C}^{\beta}$ , and  $^{13}\text{CO}$ . Red, beige, and gray colors represent helix, coil, and strand. White areas denote such strips where assignments could not be achieved.



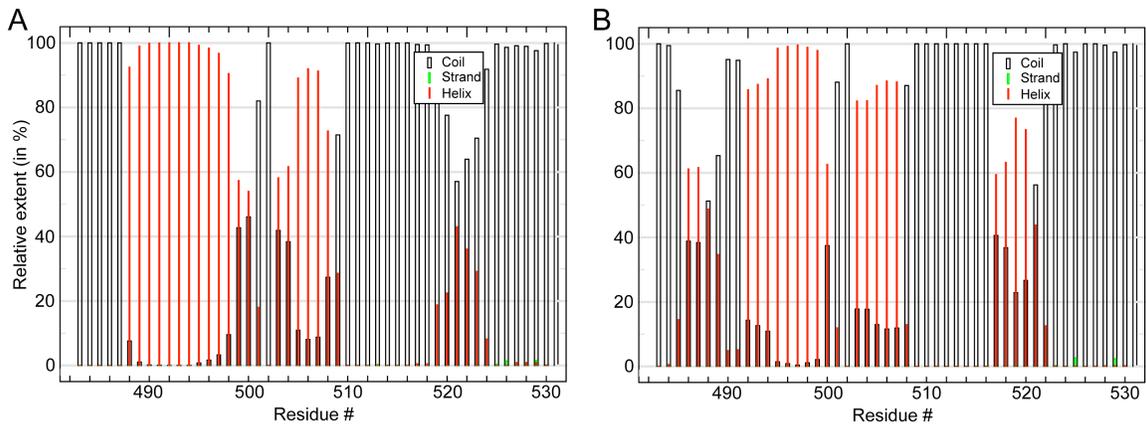
**Figure 9:** MTSL spin-labeled single-cystein mutants of the SREBP anchor as shortened by S1P. The plots represent samples with the spin label attached to S486C (**A**) and S499C (**B**), respectively, as indicated by orange arrows. Reduction of the spin label was achieved with 1 mM ascorbic acid in the same sample and buffer exchange into PBS. Scaling of the peak intensities was pursued as to reflect the signal with respect to the average peak height of observable signals (1 unit). The shaded region depicts integral membrane space as deduced from water accessibility data. Only resolved and unambiguous peaks are shown.



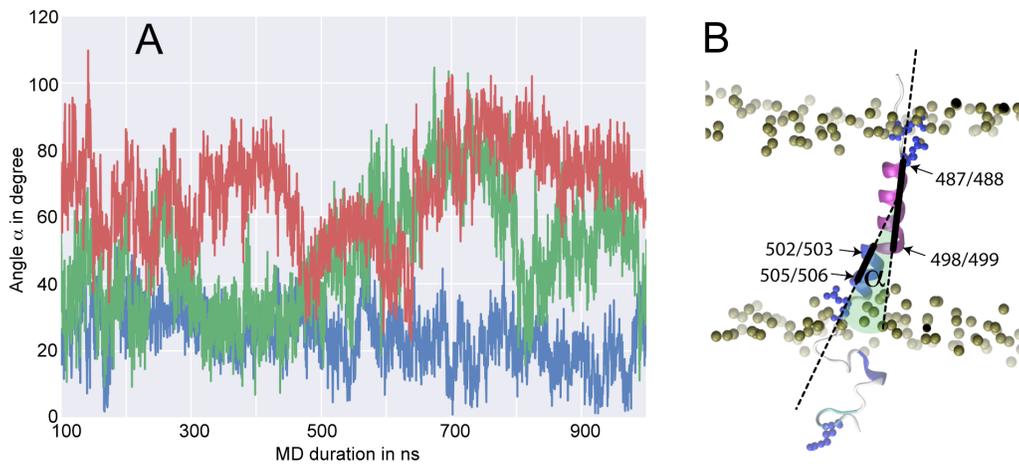
**Figure 10:** Structural models of the SREBP membrane anchor in micelles, derived from (A) TALOS+ restraints and amide-amide NOEs only, and (B) TALOS+ and NOEs restraints in wild-type protein plus PREs of two MTSL-labeled single-cysteine mutants in LMPG micelles. Shown are the lowest-energy structures of 100 structures from CNS. PREs are not able to effectively restrain the bundle due to the relative flexibility of the two helical parts to the spin label and the small size of the protein. Only the lipid-embedded part of the anchor is shown here for clarity, whereas the linker region is flexible and samples a large conformational space. Accordingly, only PREs of non-water accessible residues were considered in B). C) One of the lowest energy structures of the ensemble in micelles including flexible residues and the approximate water boundaries. (Grey and blue colors approximate the micelle and aqueous space, respectively).



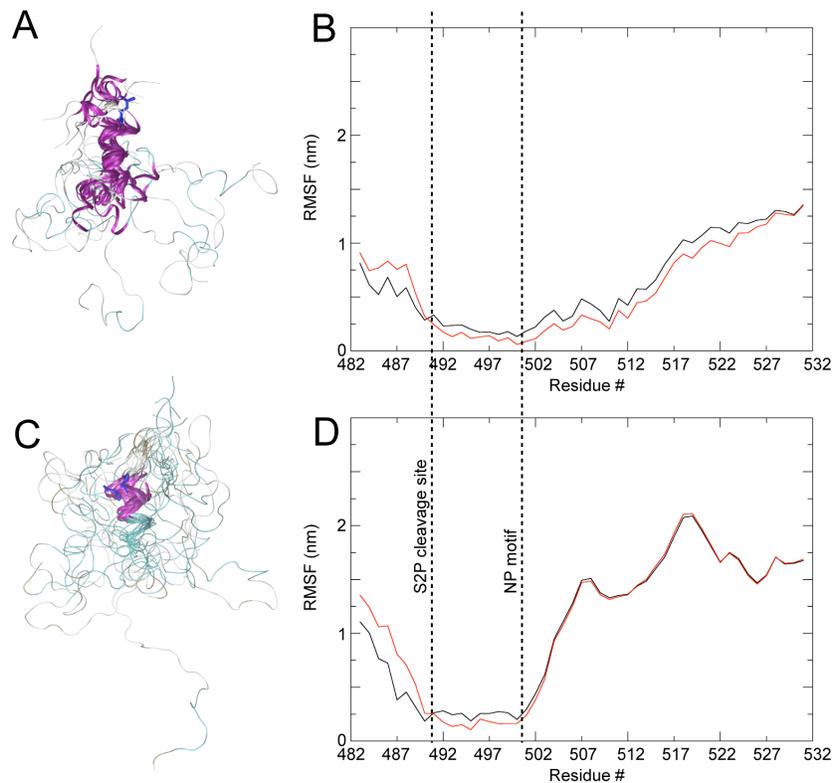
**Figure 11:** Raw data for backbone dynamics of the S1P-processed membrane anchor. **A)** Heteronuclear Overhauser effect (hetNOE), **B)** transverse relaxation rates  $R_2$ , **C)** longitudinal relaxation rates  $R_1$ , **D)** longitudinal cross relaxation ( $\eta_z$ ) and **E)** transverse cross relaxation ( $\eta_{xy}$ ) of  $^{15}\text{N}$ . See above for information about acquisition experimental details. The S1P and S2P cleavage sites are marked red, the NP motif depicted in green. The background color of the plots denotes the position of the amino acids as membrane imbedded (beige) or solvent-exposed (blue). Determination of relaxation parameters involves only dispersed peaks in all experiments. In addition, the Ser and Thr-rich solvent-exposed C terminus is largely exchange-broadened below detection in the  $T_1$  and  $T_2$  relaxation experiments.



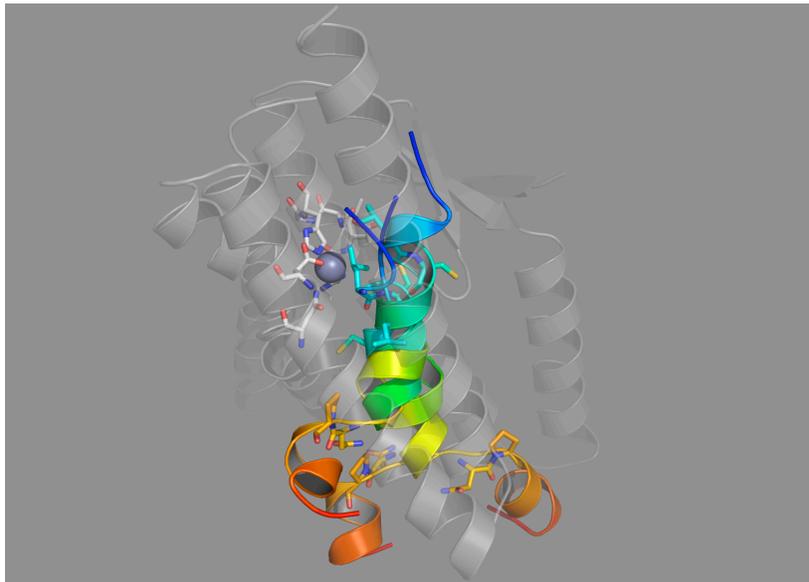
**Figure 12:** Secondary-structural stability in Molecular-Dynamics simulations as shown in Figure 4 of the Main Manuscript. A) SREBP-1 equilibrated in DLPC membranes. B) SREBP-1 equilibrated in membranes with exchange of the environment against water. Random coil,  $\beta$ -strand, and  $\alpha$ -helical structure are represented as black, green, and open red bars, respectively. The relative extent of the different types of secondary structure over time were extracted from Amber calculations over 760 and 900 ns in A) and B), respectively, with equidistant snapshots spaced by 0.4 ns.



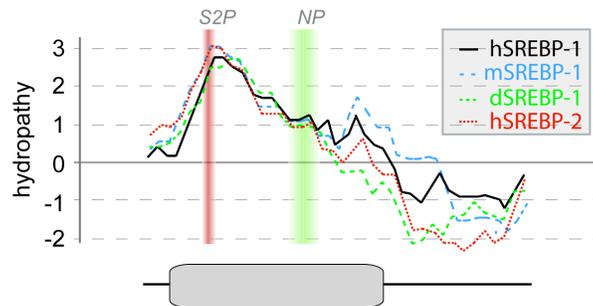
**Figure 13:** Fluctuations of the hinge-motional angle of the SREBP anchor in MD simulations using an Amber99sb-ILDN force field. The green, blue, and red curves signify the angle measured between the first and second helical part above and below the hinge motif in DLPC, POPC, and water environment, respectively. The hinge angle is defined here as the angle between two vectors in the helix direction, i. e. i) the vector connecting the mean between  $C^\alpha$  of residues 487 and 488 with the mean between  $C^\alpha$  of 498 and 499 and ii) the vector connecting the mean between  $C^\alpha$  of residues 502 and 503 with the mean between  $C^\alpha$  of residues 506 and 507.



**Figure 14:** Resilience of the SREBP-1 scissile bond to unfolding in MD simulations in a polar environment. The anchor was equilibrated in membranes with lipids subsequently exchanged against water. The calculations were run using an Amber (**A**, **B**) and a c36 tip3p force field (**C**, **D**) over 900 and 1000 ns, respectively. **A**, **C** show equidistant snapshots spaced by 0.4 ns, **B** and **D** depict root-mean-square fluctuations as a function of residue. For alignment, residues 487-502 (black curves) or 492-502 (red curves) were used.



**Figure 15:** Clusterpro (<http://cluspro.bu.edu/> protein-protein docking) docking of the NMR structure of the human SREBP-1 membrane anchor (rainbow colored) into the open conformation (PDB entry 3B4R, Chain A) of *methanocaldococcus jannaschii* S2P (gray). (26) Unstructured parts of SREBP were excluded to accelerate the docking procedure. The cleavage site of the substrate (L490-C491) is depicted as cyan sticks, the N501-P502 helix breaker motif is shown as orange sticks. Some of the docking poses (three of which are depicted) fit reasonably well within the space of the open mjS2P (i. e., orientation and distance of the cleavage site with respect to the active-site Zinc, shown as a gray sphere). At least initial binding of SREBP could involve an  $\alpha$ -helical conformation, owing to ample space in the mjS2P binding pocket. The scissile peptide bond might still be unfolded by chaperone activity of the enzyme in a second step and cleaved in an extended  $\beta$ -strand conformation, as it would be for protease cleavage in soluble proteases. These docking results (for an artificial enzyme-substrate combination), however, need to be treated with caution.



**Figure 16:** Comparison of sequence-specific hydropathy between different SREBP anchors. The characteristic tilt in the profiles is conserved among different kingdoms as well as between SREBP variants. SREBP-1 is depicted as taken from human, mouse, and drosophila, SREBP-2 is shown as the human form. The lower plot denotes the transmembrane domain.

## References:

1. Schwarz D, *et al.* (2007) Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nat. Protoc.* 2(11):2945-2957.
2. Delaglio F, *et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6:277-293.
3. Vranken WF, *et al.* (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696.
4. DeLano WL (2002) *The PyMOL Molecular Graphics System DeLano Scientific, San Carlos, CA, USA.*
5. d'Auvergne EJ & Gooley PR (2008) Optimisation of NMR dynamic models. *J. Biomol. NMR* 40(2):107-133.
6. Pelupessy P, Espallargas GM, & Bodenhausen G (2003) Symmetrical reconversion: Measuring cross-correlation rates with enhanced accuracy. *J. Magn. Reson.* 161:258-264.
7. Pelupessy P, Ferrage F, & Bodenhausen G (2007) Accurate Measurement of Longitudinal Cross-Relaxation Rates in Nuclear Magnetic Resonance. *J Chem Phys* 126:134508.
8. Shen Y, Delaglio F, Cornilescu G, & Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44(4):213-223.

9. Brünger AT, *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 54(Pt 5):905-921.
10. Hess B, Kutzner C, van der Spoel D, & Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comp.* 4(3):435-447.
11. Wolf MG, Hoefling M, Aponte-Santamaría C, Grubmüller H, & Groenhof G (2010) *g\_membed*: Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation. *J Comp Chem* 31(11):2169-2174.
12. Berendsen HJC, Postma JPM, DiNola A, & Haak JR (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684-3690.
13. Bussi G, Donadio D, & Parrinello M (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys. Lett.* 126(1):014101.
14. Essmann U, *et al.* (1995) A smooth particle mesh ewald potential. *J. Chem. Phys.* 103:8577-8592.
15. Hess B (2007) P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comp.* 4:116-122.
16. Miyamoto S & Kollman PA (1992) SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comp. Chem.* 13:952-962.
17. Lindorff-Larsen K, *et al.* (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78:1950-1958.
18. Cordoní A, Caltabiano G, & Pardo L (2012) Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters. *J. Chem. Theory Comp.* 3:948-958.
19. Jämbeck JPM & Lyubartsev AP (2012) Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids. *J. Phys. Chem. B* 10:3164-3179.
20. Joosten RP, *et al.* (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res.* 39:D411-D419.
21. Buchan DW, Minneci F, Nugent TC, Bryson K, & Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* 41:W349-357.
22. Zviling M, Kochva U, & Arkin IT (2007) How important are transmembrane helices of bitopic membrane proteins? *Biochim. Biophys. Acta* 1768(3):387-392.
23. Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105-132.
24. Rice P, Longden I, & Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16(6):276-277.
25. Cole C, Barber JD, & Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* 36(2):W197-W201.
26. Feng L, *et al.* (2007) Structure of a Site-2 Protease Family Intramembrane Metalloprotease. *Science* 318(5856):1608-1612.