Real-time conformational changes and controlled orientation of native proteins inside a protein nanoreactor

Veerle Van Meervelt[1,2], Misha Soskine[2], Shubham Singh[2], Gea K. Schuurman-Wolters[2], Hein J. Wijma[2], Bert Poolman[2] and Giovanni Maglia[2,*]

[1]Department of Chemistry, University of Leuven, Leuven, Belgium

[2]Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, the Netherlands

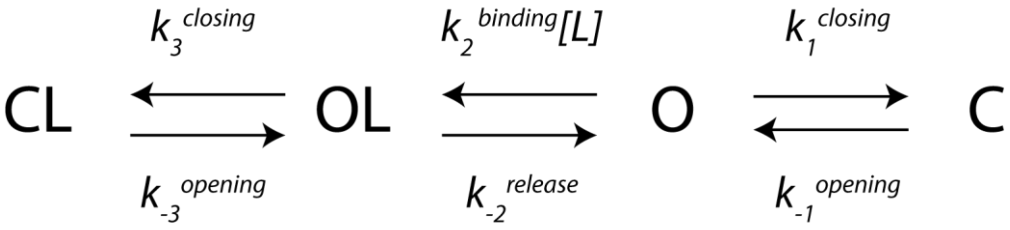
* Corresponding author: Tel. +31(0)50 363 6138; Email: g.maglia@rug.nl

Contents

Additional Results and Discussion

Mechanism for the binding of ligands to SBD1 and SBD2

The kinetic model presented in **Scheme 1** of the main text was obtained by considering the following experimental observations: (i) The SBD proteins can close with or without the ligand (L). (ii) The average opening rate decreases with substrate concentration. (iii) At saturating substrate concentrations the open conformation(s) of SBD1 and SBD2 are still observed (**Figure S6**). (iv) The $k_{closing}$ of SBD1 is not linear (**Figure 2E** in the main text). The simplest kinetic scheme that can explain all of these kinetic behaviours is described by the following.



Scheme S1. Proposed kinetic scheme for substrate binding to SBD1 and SBD2. O is the open conformation of SBD, C is the closed state without ligand, OL is the open state with ligand bound, and CL is its closed state with ligand bound.

Three of the species in **Scheme S1** were experimentally observed in this study (O, C, and CL). Notice that the closed state without ligand bound has been observed by X-ray crystallography for several substrate-binding proteins.[1–3] For species OL, it is reasonable to assume that they occur as an initial encounter complex during formation of the closed, ligand-bound state as experimentally observed by X-ray crystallography of several substrate-binding proteins[3–5].

Based on **Scheme S1**, it is possible to derive equations that describe the probability (*P*) of occurrence of each of these species. These equations (**S1-S4**) were derived under Mathematica (Wolfram Research, Inc.), using the determinant method.[6]

$$P_O = \frac{k_{-1}k_{-2}k_{-3}}{k_1k_{-2}k_{-3} + k_{-1}(k_{-2}k_{-3} + k_2(k_3 + k_{-3})\,[\mathrm{L}])} \quad \text{(S1)}$$

$$P_{OL} = \frac{k_{-1}k_2k_{-3}\,[\mathrm{L}]}{k_1k_{-2}k_{-3} + k_{-1}(k_{-2}k_{-3} + k_2(k_3 + k_{-3})\,[\mathrm{L}])} \quad \text{(S2)}$$

$$P_C = \frac{k_1k_{-2}k_{-3}}{k_1k_{-2}k_{-3} + k_{-1}(k_{-2}k_{-3} + k_2(k_3 + k_{-3})\,[\mathrm{L}])} \quad \text{(S3)}$$

$$P_{CL} = \frac{k_{-1}k_2k_3\,[\mathrm{L}]}{k_1k_{-2}k_{-3} + k_{-1}(k_{-2}k_{-3} + k_2(k_3 + k_{-3})\,[\mathrm{L}])} \quad \text{(S4)}$$

Thus, considering a single enzyme, the observed opening rate $k_{opening}$ is described by the fraction of molecules that reach the C configuration $\frac{P_C}{P_C + P_{CL}}$ and the C.L configuration $\frac{P_{CL}}{P_C + P_{CL}}$ as:

$$k_{\text{opening}} = \frac{k_{-1}\,P_C}{P_C + P_{CL}} + \frac{k_{-3}\,P_{CL}}{P_C + P_{CL}} \quad \text{(S5A)}$$

After substitution of **equations S1** to **S4** into **S5A**, the result can be simplified to:

$$k_{\text{opening}} = \frac{k_{-1}k_{-3}\,(k_1k_{-2} + k_2k_3\,[\mathrm{L}])}{k_1k_{-2}k_{-3} + k_{-1}k_2k_3\,[\mathrm{L}]} \quad \text{(S5B)}$$

This **equation S5B** can be rearranged to:

$$k_{opening} = k_{-3} + \frac{k_1 k_{-1} k_{-2} k_{-3} - k_1 k_{-2} k_{-3}^{\,2}}{k_1 k_{-2} k_{-3} + k_{-1} k_2 k_3 [\mathrm{L}]} \qquad \text{(S5C)}$$

Equation S5C demonstrates that at high ligand concentration $k_{opening}$ should approach k_{-3}, because the right term vanishes due to the [L] term in the denominator. Furthermore, by filling in a ligand concentration of zero, **equation S5C** simplifies to:

$$k_{opening} = k_{-1} \quad \text{(if [L] = 0)} \qquad \text{(S5D)}$$

Thus, the experimentally observed reduction in the observed $k_{opening}$ with increasing [L] indicates that $k_{-1} > k_{-3}$ which is in agreement with ligand binding stabilizing the closed state.

With the same rationale used to derive equations for the closing rate, the observed $k_{closing}$ is:

$$k_{closing} = \frac{k_1 P_O}{P_O + P_{OL}} + \frac{k_3 P_{OL}}{P_O + P_{OL}} \qquad \text{(S6A)}$$

After substitution **equations S1-4**, this simplifies to:

$$k_{closing} = \frac{k_1 k_{-2} + k_2 k_3 [\mathrm{L}]}{k_{-2} + k_2 [\mathrm{L}]} \qquad \text{(S6B)}$$

This can be rewritten as:

$$k_{\text{closing}} = k_3 + \frac{k_1 k_{-2} - k_{-2} k_3}{k_{-2} + k_2[\text{L}]} \qquad \text{(S6C)}$$

At very high ligand concentration the rate should approach k_3. This can explain the observed leveling off in opening rate of SBD1 at saturating concentration of ligand (**Figure 2E**).

At zero ligand concentration, one obtains

$$k_{\text{closing}} = k_1 \quad \text{(if [L] = 0)} \qquad \text{(S6D)}$$

Thus, the intercept of k_{closing} versus [L] should be higher than zero, which is indeed observed (**Figure 2E** and **3D** in main manuscript).

Using the same rationale, it is possible to predict the fraction of the closed state f_{closed} versus ligand concentration (**equations S7**).

$$f_{\text{closed}} = \frac{P_C + P_{CL}}{P_C + P_{CL} + P_O + P_{OL}} \qquad \text{(S7A)}$$

After substituting **equations S1-S4**, **equation S7A** can be simplified to:

$$f_{\text{closed}} = \frac{k_1 k_{-2} k_{-3} + k_{-1} k_2 k_3\,[\text{L}]}{k_1 k_{-2} k_{-3} + k_{-1}(k_{-2} k_{-3} + k_2(k_3 + k_{-3})\,[\text{L}])} \qquad \text{(S7B)}$$

To obtain estimates for the kinetic constants in **Scheme S1**, a simultaneous fit of all the experimental data to **equations S5B**, **S6B**, and **S7B** was carried out. Least-squares fitting was performed using the solver function in Excel (Microsoft) as described elsewhere.[7] During the least-squares fit, 1/SD was used as weight, in which SD is the standard deviation in the measurements for each data point. The fittings were in agreement with the experimentally observed result (**Figure 2E and 3D**), with the resulting rate constants shown in **Table S3 and S4**. The overall agreement between experimentally observed and simulated kinetic values was satisfactory. All aspects of the kinetic behavior of SBD1 and SBD2 were well reproduced.

The 95% confidence intervals were determined with the Fisher's F distribution method.[7] In this method, the value of het sum of squared residuals (SSR) that belongs to the confidence interval is calculated considering the number of parameters that are fit, the number of data points, and the value of SSR that was obtained when all parameters were fitted to an optimal value. The fit to the data is subsequently optimized while fixing one of the parameters at a series of values that increasingly deviate from optimal. The value at which this parameter forces the SSR to exceed the threshold value is the border of a 95% confidence interval. This procedure is done separately for all fitted parameters or combinations of these.

Ligand-induced binding of glutamine to SBD2 variants

Several SBD2 variant were prepared to control the orientation of the protein inside the nanopore. Although the studied SBD2 variants were carefully selected by MD simulations not to alter the structural stability of the protein, we observe that the introduction of a charged residue into SBD2 did influence the protein dynamics. In the

absence of substrate, all tested SBD2 variants still showed intrinsic closing to level L_C. The variants that favored L_{OA} blockades (orientation A) showed an intrinsic opening rate of 29.7 ± 3.8 s^{-1}, 32.5 ± 7.4 s^{-1} and 32.6 ± 2.1 s^{-1} for T256E, S358K and T256E+S358K, respectively, which is lower than the intrinsic opening rate observed with wildtype SBD2 (55.6 ± 3.1 s^{-1}). In contrast, the variants that favored L_{OB} blockades (orientation B) showed a higher (or similar) intrinsic opening rate (91.6 ± 6.1 s^{-1}, 55.6 ± 4.1 s^{-1} and 85.5 ± 16.8 s^{-1} for T256K, S358E and T256K+S358E, respectively, **Figure S4B**). All tested variants could bind glutamine, although the SBD2 variants that adopt orientation B showed a reduced affinity as shown by the reduced fraction of bound state at 1 µM glutamine (**Figure S4C**). The mutants that favored orientation B showed faster opening rates (**Figure S4D**, **Table S1**) and lower closing rates than the wild type protein, except for S358E (**Figure S4**). Therefore, our results suggest that the substitutions only have a small effect on the structural stability of the SBD2 variants. Most interestingly, two different bound state levels were observed for SBD2(T256E) and SBD2(S358K) variants (**Figure S3**), suggesting that the bound state can also dwell inside the nanopore with two orientations. The variant that combines both substitutions SBD2(T256E+S358K) adopts only orientation B, *i.e.* it only shows a single current level for the closed conformation.

Materials and methods

All chemicals were purchased from Sigma-Aldrich, enzymes from Thermo Scientific and DNA from Integrated DNA Technologies (IDT) unless otherwise specified. Throughout the text, '*N*' indicates the number of independent nanopore experiments and '*n*' the total number of data points in the data set. All errors in this work are given as standard deviations unless otherwise specified.

Modelling of SBD2 variants

Bioinformatics and molecular modelling were performed to find amino acid substitutions that would alter the orientation of SBD2 inside the pore, that is, by redistributing surface charges without decreasing the stability of SBD2. Only mutations to aspartate (D), glutamate (E), and lysine (K) were considered. Histidine residues were excluded as its pK_a, which is typically between 5 and 8[8], is close to the pH of the assay buffer. Arginines were avoided, because they are more likely to cause solubility problems.[9] Sequence alignment was used to find positions in homologs of SBD2 where D/E/K residues are present. Such occurrences in proteins with sufficient similarity suggests that the corresponding mutations will be tolerated at these positions[10]. CS-blast[11] was used to search for proteins with at least 35% identity. The collected sequences were aligned with MUSCLE[12,13]. At any given position, only those D/E/K mutations that occurred in > 5% of the homologous proteins were further considered. This totalled to 133 potential mutations at 84 different positions (see **Table S5**). Conservation scores (**Figure 4C**) were assigned using ConSurf.[14,15]

The Rosetta supercharging application was subsequently used because it is specifically optimized for the task of altering the charge of a protein by introducing mutations that are compatible with folding.[16] The command line options used for Rosetta supercharging were -use_input_sc -ignore_unrecognized_res -surface_residue_cutoff 16 -dont_mutate_glyprocys true -dont_mutate_correct_charge true -dont_mutate_hbonded_sidechains true -include_lys -include_asp -include_glu -refweight_arg -1.98 -refweight_lys -1.65 -refweight_asp -0.6 -refweight_glu -0.8 -target_net_charge_active true -resfile resfile_5percent -target_net_charge X. In these options, resfile_5percent was a file listing the potential mutations obtained from bioinformatics (**Table S5**) and X indicated the requested net charge. The calculations were done repeatedly, each time shifting X with a unit. Charges from -30 to +30 were requested (the net charge of the native protein is -4). Mutations that were not prioritized by Rosetta supercharging were eliminated. This resulted in selection of 14 mutations that introduced a lysine (D279K, Q335K, S345K, S354K, S358K, N383K, A384K, E388K, T392K, S451K, A468K, D473K, D477K, E481K) and 18 mutations (T256E, Q276E, K302E, A332D, S345E, S358E, K363D, K366E, N383E, A384D, K387D, T392E, T400D, K430D, K430E, K436E, S451E, N461D) that introduced a negative charge.

Molecular dynamics (MD) simulations were carried out to verify that the selected mutations did not lead to significant increases in flexibility, which would indicate lower structural stability.[17] A simulation protocol under YASARA[18,19] was applied that had already been demonstrated to successfully eliminate destabilized mutants based on their increased flexibility[20]. This protocol is described in detail in previous work and all scripts for simulation and analysis are available upon request.[20,21] Its main feature is that for each modelled structure five independent MD simulations are performed,

which are started with different initial atom velocities. The use of such parallel independent MD simulations results in better sampling of protein conformations than a single longer MD simulation.[22–24] A total of 14 point mutations (T256E, Q276E, K302E, A332D, S345E, S354K, S358K, E383K, N383K, A384K, E388K, T392K, S451E, N461D) that did not increase protein flexibility were further considered for experimental characterization.

From these, mutations at position T256 and S358 (T256K/E and S358K/E) were prepared because they showed the lowest conservation and were the farthest apart from each other. For position 256 no negative charge introducing mutation had been selected. The mutation T256E had earlier been eliminated because Rosetta did not prioritize it. However, additional MD simulations showed that T256E did not result in increased flexibility, suggesting it would be compatible with the structure.

To further ensure that these potential mutations were compatible with the protein structure, they were modelled by FoldX. Like Rosetta, FoldX uses an energy function that balances the terms for van der Waals interactions, geometric preferences of angles and dihedrals, entropy, *etcetera* to model the 3D structure of mutant variants.[25] Rosetta and FoldX[26] have different energy functions, which should make the calculation outcomes orthogonal and increase the chance that the selected mutations are indeed tolerated. Also, by subtracting the energies calculated for the native structure from those of the mutant structure, FoldX predicts the change in the free energy of folding due to the mutation ($\Delta\Delta G^{fold}$). The command line options used for FoldX were: --command=BuildModel --mutant-file=individual_list.txt --numberOfRuns=5. These settings make FoldX use its standard protocol for introducing the mutations listed in the file individual_list.txt and produce a predicted $\Delta\Delta G^{fold}$ that was obtained by averaging the results from five calculation runs per mutant. This

protocol predicted that the changes in $\Delta\Delta G^{fold}$ were smaller than the prediction error of the method (3.4 kJ/mol[25]) for all four mutations that were finally selected for experimental characterization (T256E, -0.1 kJ/mol; T256K, -1.1 kJ/mol; S358E, 2.0 kJ/mol; S358K, 0.9 kJ/mol).

Preparation of SBD2 variants

SBD2 variants were prepared according to the MEGAWHOP procedure.[27] Template DNA (pBADnLIC containing SBD2) was amplified using primers listed in **Table S6**. REDTaq® ReadyMix™ (50% (v/v)) was mixed with 0.5 µM of forward and reverse primers and ~100 ng of plasmid template in a final volume of 300 µL (six reactions). After a pre-incubation step at 95ºC for 3 minutes, the reactions were cycled 27 times according to the cycling protocol: denaturation at 95ºC for 15 s, annealing at 55ºC for 15 s, extension at 72ºC for 2 min. The resulting PCR products were pooled, purified and concentrated using the QIAquick PCR purification kit and run on an agarose gel (0.8%). The PCR product was then gel extracted using the QIAquick Gel Extraction kit. Next, ~500 ng of the purified PCR product was mixed with ~50 ng of the template plasmid and the amplification was carried out with Phire Hot Start II DNA polymerase. After 30 s pre-incubation at 98ºC, the reaction was cycled 30 times according to the following cycling protocol: denaturation at 98ºC for 5 s, extension at 72ºC for 1.5 min. The circular template was eliminated by incubation with Dpn I (1 FDU) for 1 hour at 37ºC. The resulted mixture was transformed into E. cloni® 10G cells (Lucigen) by electroporation. The transformed bacteria were grown overnight at 37ºC on ampicillin (100 µg/ml) LB agar plates. The identity of the clones was confirmed by sequencing.

Purification of SBD1, SBD2 and variants

Substrate binding domains carrying a histidine-tag were expressed in *E. coli* strain MC1061 and purified as described previously.[28] Cell lysate was mixed with 50 mM potassium phosphate (KP_i), pH 8.0, 200 mM KCl, 20% glycerol (buffer A) supplied with 20 mM imidazole and incubated with Ni^{2+}-Sepharose resin for 1h at 4°C. Next, the resin was washed with 20 column volumes of buffer A supplied with 50 mM imidazole, followed by protein elution with 500 mM imidazole in buffer A. To prevent protein aggregation, 5 mM EDTA was added immediately after elution. Afterwards imidazole was removed by buffer exchange to 50 mM Tris-Cl, 0.5 mM DTT, 0.5 mM EDTA, pH 8.0, and the histidine-tag was cleaved off using His-TEV(S219V) by overnight incubation.[29] The histidine-tagged TEV protease and residual uncut protein were removed by incubating the mixture with Ni-NTA for 1h and collecting the flow-through. Next, proteins were further purified with size-exclusion chromatography on a Superdex-200 column (GE Healthcare) equilibrated with 50 mM Tris-HCl, 200 mM NaCl, pH 7.5. Peak fractions were collected and supplemented with 25% glycerol and aliquots were stored at -80°C after flash freezing in liquid nitrogen.

Expression and purification of ClyA nanopores

In this work we used an engineered variant of ClyA from *Salmonella typhi*, ClyA-AS (C87A/L99Q/E103G/F166Y/I203V/C285S/K294R/H307Y). Both variants contain a C-terminal hexa-histidine tag. ClyA-AS monomers were expressed in E. cloni® EXPRESS BL21 (DE3) cells and purified using Ni-NTA affinity chromatography. Oligomers were then formed by adding 0.2% *n*-Dodecyl-β-D-Maltoside (DDM, GLYCON Biochemicals GmbH). Type I ClyA-AS oligomers were separated from

monomers and several other oligomeric ClyA forms[30], using Blue-Native poly-acrylamide gel electrophoresis (BN-PAGE, Bio-Rad). Oligomeric ClyA proteins show multiple bands in Blue-Native gels. In this work we extracted the lowest band from the poly-acrylamide gel, which is likely to correspond to the 12-meric (Type I) form of ClyA. Aliquots were stored at 4°C in 150 mM NaCl, 15 mM Tris.HCl, pH 7.5 supplemented with 0.2% DDM and 10 mM EDTA.[31]

Electrical recordings in planar lipid bilayers

Single channel recordings were performed as described before.[32] In short, a 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids) lipid bilayer was formed across a ~100 µm Teflon (Goodfellow Cambridge Limited) aperture separating two compartments (*cis* and *trans*) filled with 150 mM NaCl, 15 mM Tris-HCl, pH 7.5. An electrical potential was applied to the *trans* side using Ag/AgCl electrodes. ClyA nanopores were inserted into lipid bilayers from the *cis* compartment, which was connected to the ground. After the insertion of a single nanopore, the excess of protein was removed by several cycles of perfusion. SBD1 or SBD2 proteins were added to the *cis* chamber at a final concentration of 74 nM or 72 nM, respectively. Substrates were added *cis* and incubated in the chamber 2 minutes prior to recording.

Data were recorded by applying a 2 kHz low-pass Bessel filter and using a 10 kHz sampling rate. An additional digital Gaussian filter at 100 Hz was applied post-acquisition to the current recordings. The electrical signals were amplified by using an Axopatch 200B patch clamp amplifier (Axon Instruments) and digitized with a Digidata 1320 A/D converter (Axon Instruments). Data were recorded by using the Clampex 10.5 software (Molecular Devices) and subsequent analysis was carried out with

Clampfit software (Molecular Devices) and OriginPro 9.0 (OriginLab). All experiments were carried out at 24 ºC.

Determination of the kinetic parameters of SBD1 and SBD2

SBD1 experiments were performed at -60 mV and SBD2 experiments at -100mV. Residual current values ($I_{res\%}$) were calculated from blocked pore current values (I_B) and open pore current values (I_O) as $I_{res\%} = 100 * I_B / I_O$. I_O and I_B were determined from the Gaussian fit to all point current histograms. The equilibrium dissociation constant (K_d) of the SBD:substrate complex was determined from titration experiments. The SBD1 concentration in the *cis* compartment was 74 nM and asparagine was added *cis* from a stock solution (50 µM) at final concentrations of 0.2µM, 0.4µM, 0.6µM, 0.8µM, 1.0µM, 1.9µM, 2.9µM and 3.7µM. The SBD2 concentration was 72 nM and glutamine was added (from a 50 µM stock solution) at a final concentration of 0.2 µM, 0.4 µM, 0.8 µM, 1.2 µM and 2.3 µM; and a 5 mM stock solution was used for a final concentration of 10.0 µM and 19.9 µM. After obtaining the whole current histogram, K_d values were calculated from the ratio of areas under the peak of the open or closed ligand-bound population, according to this relationship: relative closed population = Closed/(Open+Closed). The plot of the relative closed SBD population *versus* the added substrate concentration gave a saturation curve that was fitted according to the one-site binding isotherm. The K_d is the concentration of substrate at 50% signal saturation.

The lifetimes of the SBD open (τ_{on}) and closed states (τ_{off}) were determined at -60mV (SBD1), or -100mV (SBD2 and variants) by using the “single-channel search” function in Clampfit. A threshold level was set for the blockades. Events shorter than 5 ms were ignored in the analysis of SBD1, while events shorter than 2 ms were ignored in the

analysis of SBD2. The process of event collection was monitored manually. The resulting event lifetimes (at least 200 events) from the open or closed state were binned together as cumulative distributions and fitted to a single exponential to retrieve the lifetimes (τ). Transition rates were calculated from the inverse of the lifetimes for each experiment and the average was calculated and plotted versus the substrate concentration. Rate constants were fitted to the kinetic schemes described in the additional discussion.

Autocorrelation analysis of SBD1 rates

Based on the theory of fluorescence correlation spectroscopy, autocorrelation analysis can be used to directly extract kinetic rates from raw dynamic traces.[33,34] First, the acquired electrical trace is divided in blocks in which one protein is continuously present in the pore. Subsequently, each block is independently base-lined by subtracting the level of the lower state (usually L_O), which is extracted by the software after applying a frequency cut-off filter from the signal. Next, the unfiltered and baselined signal $I(t)$ is correlated with itself at delay time τ, and normalized to yield the autocorrelation $G(\tau)$ (the brackets denote the discrete time average):

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2}.$$

The discrete correlation $\langle I(t)I(t+\tau)\rangle$ is calculated using a multiple-tau correlation algorithm – which can be implemented efficiently in software – for delay times

$$\tau_n = \Delta t \begin{Bmatrix} 0, 1, 2, 3, \\ 4\cdot 2^0, 5\cdot 2^0, 6\cdot 2^0, 7\cdot 2^0, \\ 4\cdot 2^1, 5\cdot 2^1, 6\cdot 2^1, 7\cdot 2^1, \\ \dots \end{Bmatrix},$$

where Δt is the sampling time of the recording[35]. The maximum delay time is limited by the finite length of the signal. For a two-state signal that stochastically changes

between an off state (zero signal) and an on state (arbitrary level) with rates k_{on} and k_{off}, the autocorrelation function equals:

$$G(\tau) = 1 + \frac{k_{off}}{k_{on}} \exp\left(-\tau\left(k_{on} + k_{off}\right)\right).$$

The rate constants can be extracted by fitting the above equation to the calculated autocorrelation function, ignoring the first two data points at $\tau = 0$ (shot noise) and $\tau = \Delta t$ (undefined noise component). For each condition, 10 blocks were analyzed and the mean and standard deviation of the kinetic rates are plotted (**Figure S7**).

Supplementary Information Figures

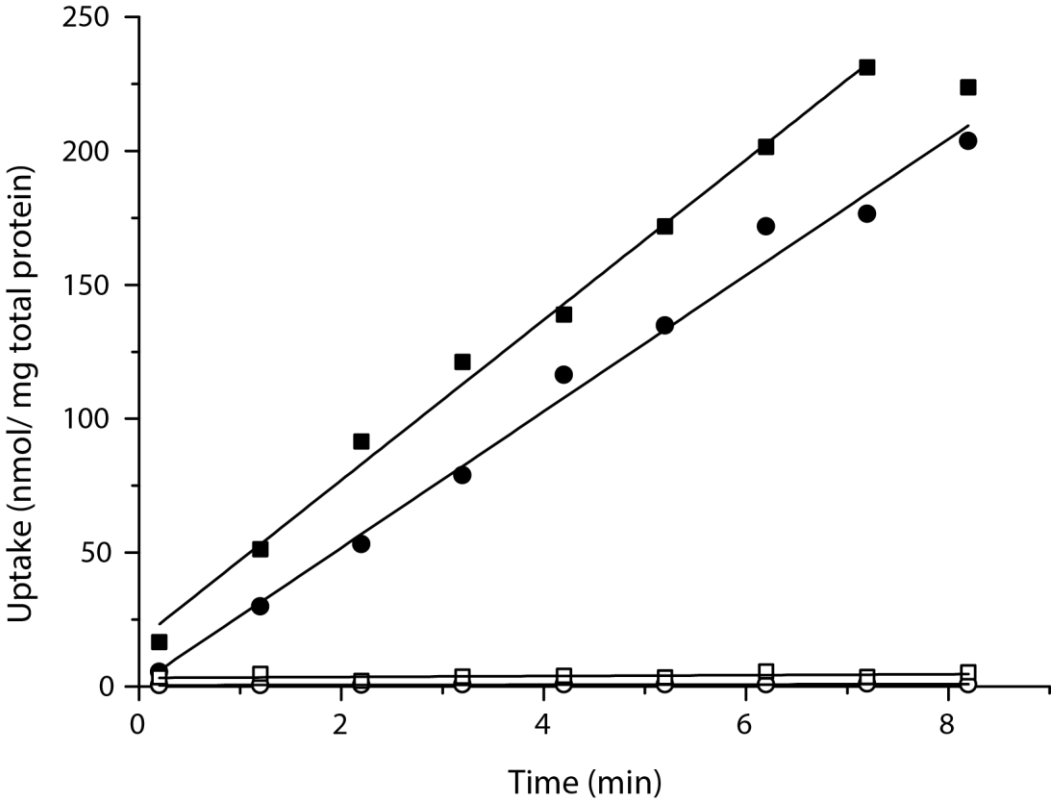


Figure S1. Amino acid uptake in *L. lactis* strain GWK9000.

Amino acid transport by GlnPQ(SBD1) (closed symbols) and GlnPQ(SBD1)E184W (open symbols) at 100 µM asparagine (circles) or 100 µM glutamine (squares). Experiments were performed as described previously.[28]

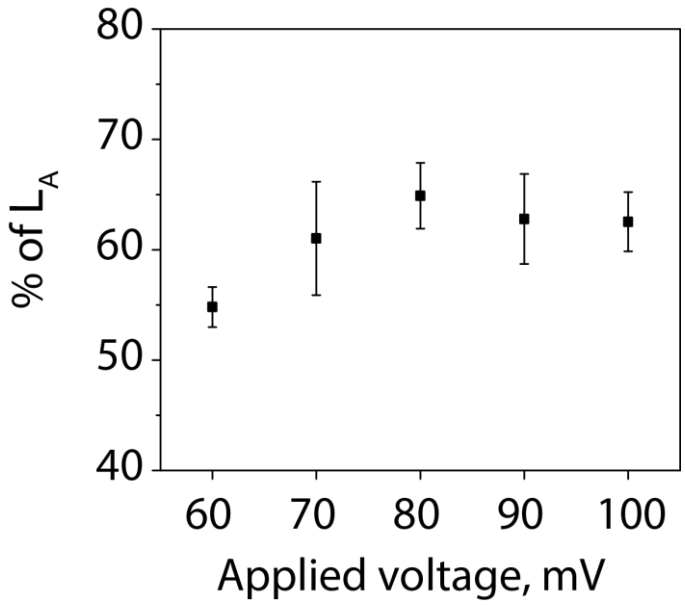


Figure S2. Voltage dependency of SBD2 L_{OA} occupancy.

Percent of L_A [L_{OA}/(L_{OA} + L_{OB})*100] sampled with Type I ClyA-AS as a function of the applied bias.

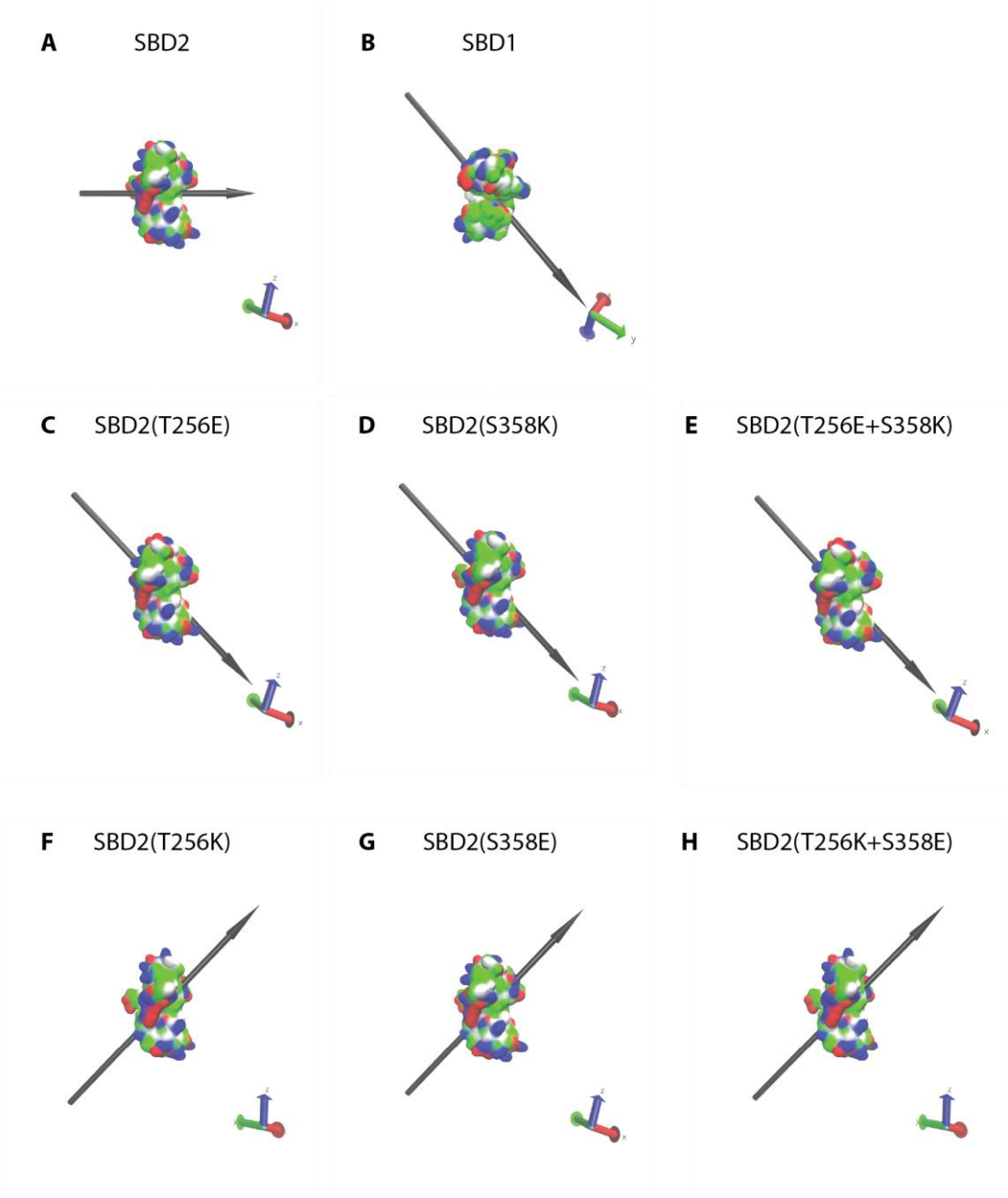


Figure S3. Dipole moments of SBD1, SBD2 and variants.

Graphic representation of the dipole moments of SBD1 (A), SBD2 (B), SBD2(T256E) (C), SBD2(S358K) (D), SBD2(T256E+S358K) (E), SBD2(T256K) (F), SBD2(S358E) (G), SBD2(T256K+S358E) (H). Dipole moments were calculated and visualized using the dipole watcher plugin of VMD[36]. Proteins are coloured according to their residue

type: basic residues are coloured blue, acidic residues red, polar residues green and nonpolar residues white.

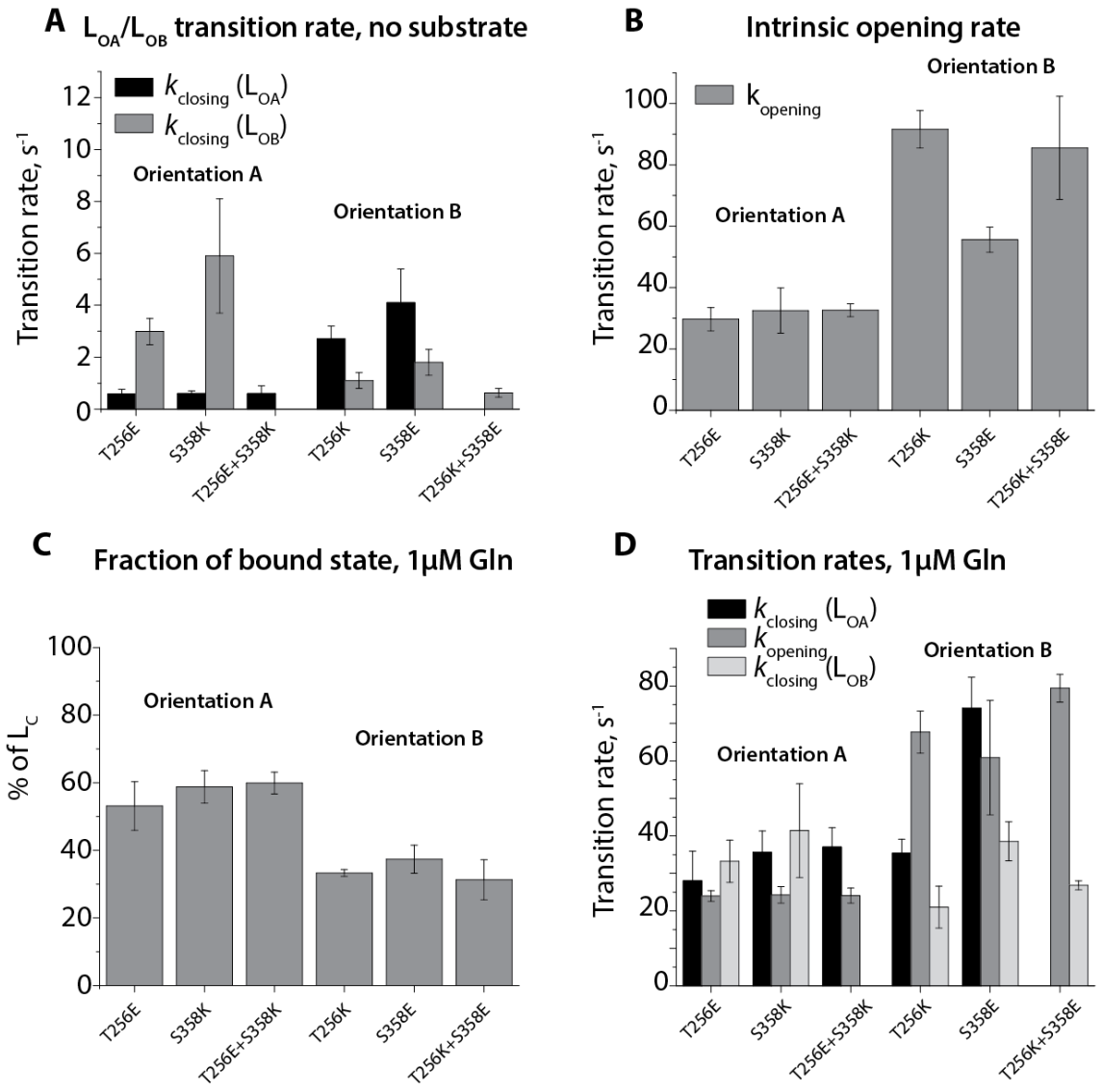


Figure S4. Transition rates of the SBD2 variants.

(A) Transition rate from L_{OA} and L_{OB} in the absence of substrate for the SBD2 variants. (B) Opening rate of the intrinsically closed conformation for the six SBD2 variants in the absence of substrate. (C) Percentage of bound state in presence of 1 µM glutamine, calculated from the ratio of areas ($L_C/(L_{OA}+L_{OB}+L_C)$) from the whole current histogram. (D) Transition rates for the SBD2 variants from L_{OA} to L_C [$k_{closing}$ (L_{OA})] and from L_{OB} to L_C [$k_{closing}$ (L_{OB})] in presence of 1 µM glutamine. All results were obtained at -100 mV in 150mM NaCl, 15mM Tris-HCl, pH 7.5.

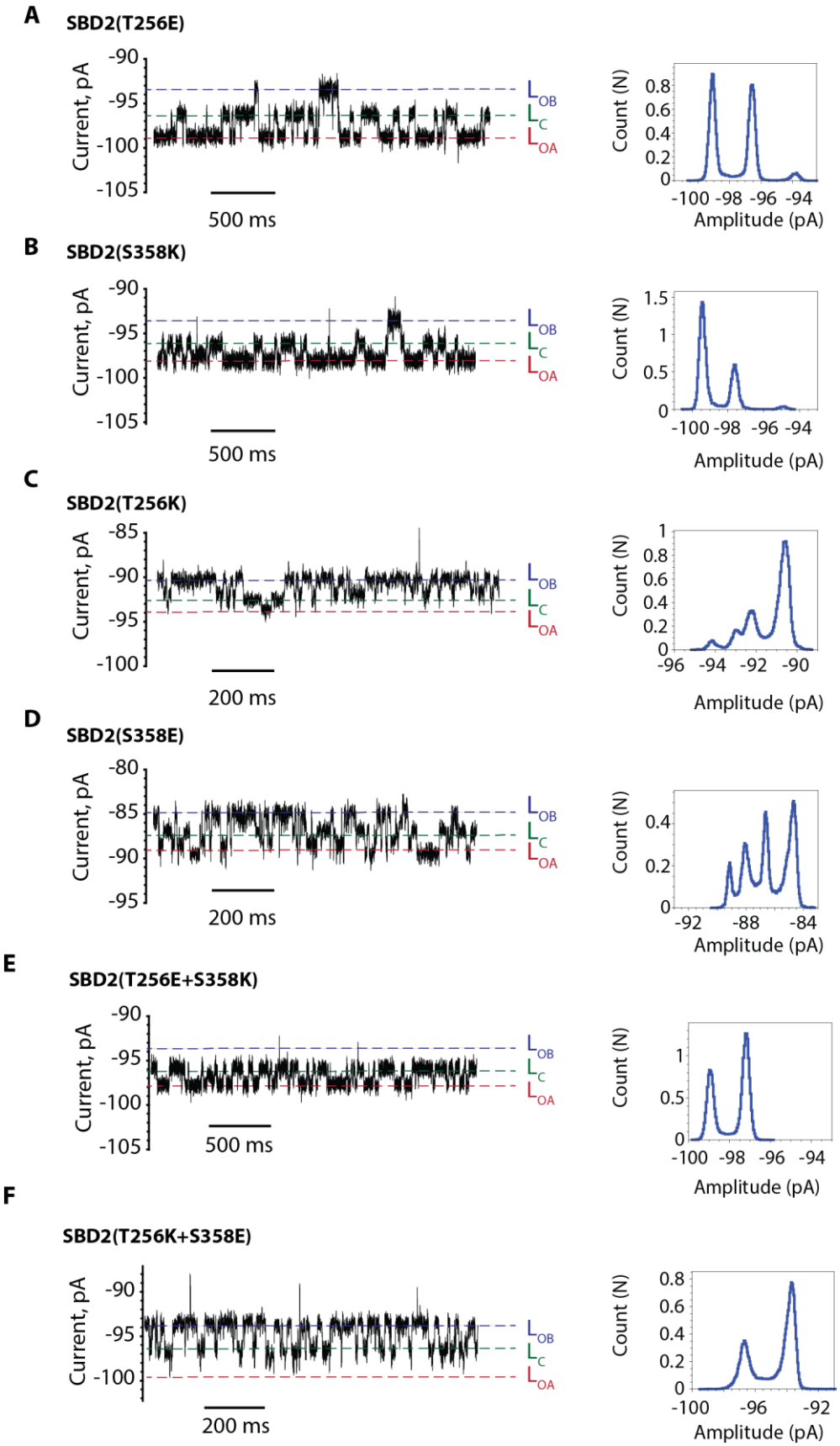

A
SBD2(T256E)
Current, pA
-90
-95
-100
-105
500 ms
L_{OB}
L_C
L_{OA}
Count (N)
0.8
0.6
0.4
0.2
0
-100 -98 -96 -94
Amplitude (pA)
B
SBD2(S358K)
Current, pA
-90
-95
-100
-105
500 ms
L_{OB}
L_C
L_{OA}
Count (N)
1.5
1
0.5
0
-100 -98 -96 -94
Amplitude (pA)
C
SBD2(T256K)
Current, pA
-85
-90
-95
-100
200 ms
L_{OB}
L_C
L_{OA}
Count (N)
1
0.8
0.6
0.4
0.2
0
-96 -94 -92 -90
Amplitude (pA)
D
SBD2(S358E)
Current, pA
-80
-85
-90
-95
200 ms
L_{OB}
L_C
L_{OA}
Count (N)
0.4
0.2
0
-92 -88 -84
Amplitude (pA)
E
SBD2(T256E+S358K)
Current, pA
-90
-95
-100
-105
500 ms
L_{OB}
L_C
L_{OA}
Count (N)
1
0.5
0
-100 -98 -96 -94
Amplitude (pA)
F
SBD2(T256K+S358E)
Current, pA
-90
-95
-100
200 ms
L_{OB}
L_C
L_{OA}
Count (N)
0.8
0.6
0.4
0.2
0
-100 -96 -92
Amplitude (pA)

Figure S5. Substrate binding to SBD2 variants.

Current blockades to Type I ClyA-AS for SBD2 variants T256E (A), S358K (B), T256K (C), S358E (D), T256E+S358K (E) and T256K+S358E (F) in the presence of 1 µM glutamine at -100 mV in 150 mM NaCl, 15 mM Tris-HCl, pH 7.5. Current traces were collected at 24°C by applying a Bessel low-pass filter with a 2 kHz cut-off and sampled at 10 kHz, a post-acquisition Gaussian filter of 100 Hz was applied.

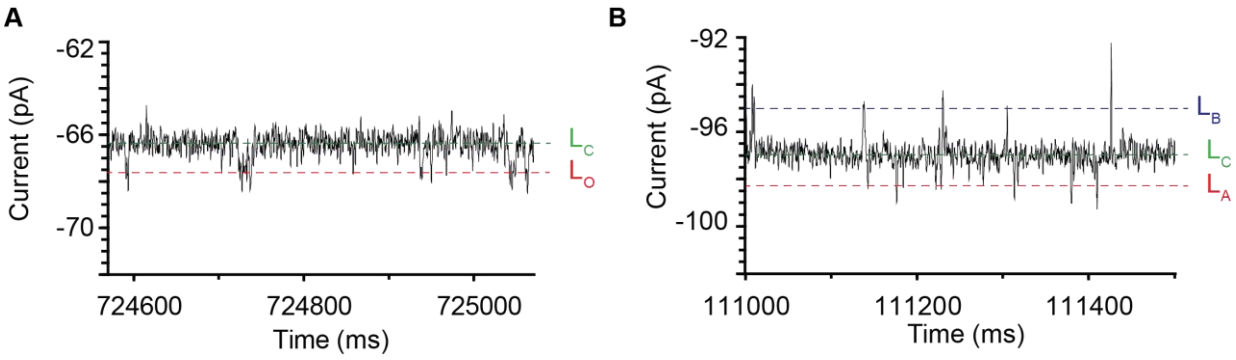


Figure S6. Details of SBD blockades at saturating substrate concentrations.

A) SBD1 current blockade in the presence of 50 µM asparagine at -60 mV. B) SBD2 current blockade in ClyA-AS in the presence of 50 µM glutamine at -100 mV. Current traces were collected in 150 mM NaCl, 15 mM Tris-HCl, pH 7.5 at 24°C by applying a Bessel low-pass filter with a 2 kHz cut-off and sampled at 10 kHz, a post-acquisition Gaussian filter of 400 Hz was applied.

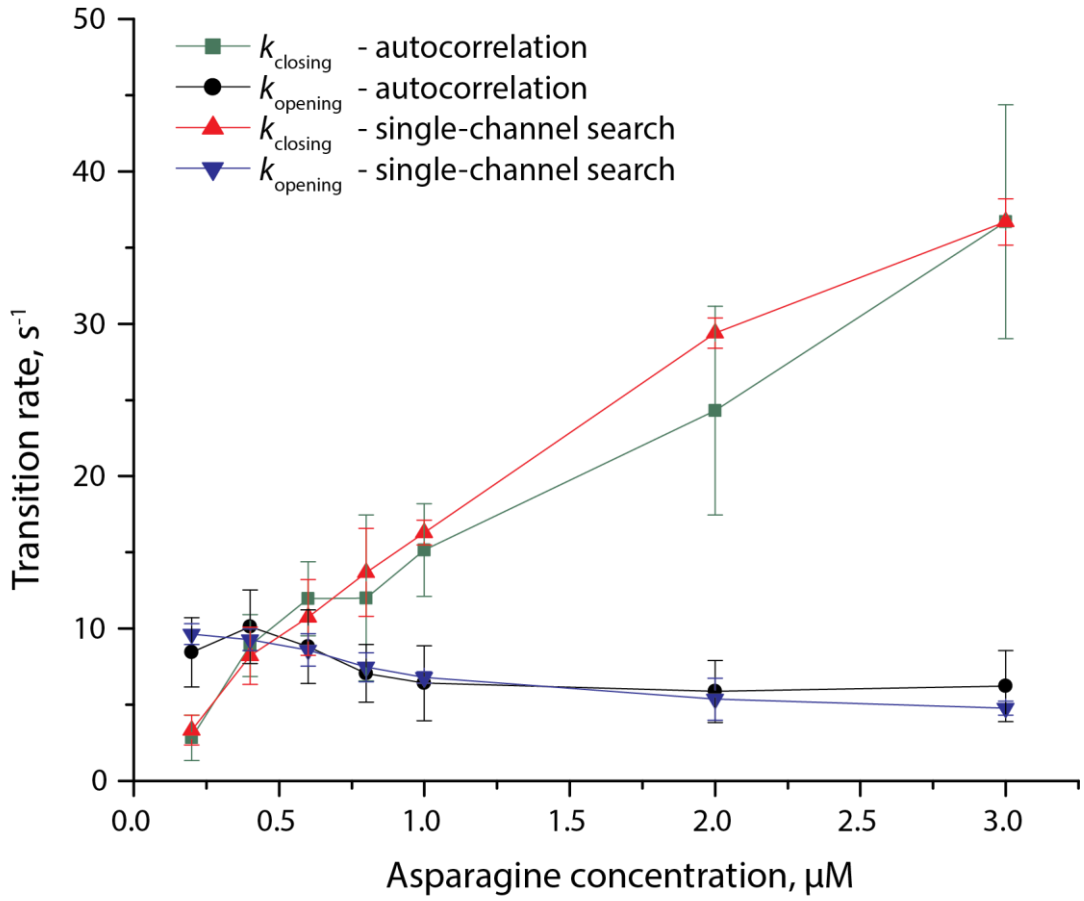


Figure S7: Comparison of SBD1 binding rates analyzed by autocorrelation and single-channel search.

Opening and closing rates were calculated from the inverse of the lifetimes analyzed using the single-channel search function of clampfit (red triangles: $k_{closing}$, blue triangles: $k_{opening}$) after a post-acquisition Gaussian filter of 100 Hz was applied; or by autocorrelation analysis (green squares: $k_{closing}$, black circles: $k_{opening}$) with no additional flitering.

Supplementary information tables

Table S1. Kinetic parameters for the SBD2 variants in the absence and presence of 1 μM glutamine.

		no substrate			1 μM Gln		
		L_{OA} to L_{OB} rate, s^{-1}	L_{OB} to L_{OA} rate, s^{-1}	L_C $k_{opening}$, s^{-1}	L_{OA} $k_{closing}$, μM^{-1} s^{-1}	L_{OB} $k_{closing}$, μM^{-1} s^{-1}	L_C $k_{opening}$, μM^{-1} s^{-1}
Orientation A	**SBD2**	3.6 ± 0.4	6.1 ± 1.3	56 ± 3	45 ± 5	49 ± 3	38 ± 4
	T256E	0.6 ± 0.2	3.0 ± 0.5	30 ± 4	28 ± 8	33 ± 6	24 ± 1.4
	S358K	0.6 ± 0.1	5.9 ± 2.2	33 ± 7	36 ± 6	41 ± 13	24 ± 2
	T256E+S358K	0.6 ± 0.1	NA	33 ± 2	37 ± 5	NA	24 ± 2
Orientation B	**T256K**	2.7 ± 0.5	1.1 ± 0.3	92 ± 6	35 ± 4	21 ± 6	68 ± 6
	S358E	4.1 ± 1.3	1.8 ± 0.5	56 ± 4	74 ± 8	39 ± 5	61 ± 15
	T256K+S358E	NA	0.6 ± 0.2	85 ± 16	NA	27 ± 1	79 ± 4

Table S2. Residual current values (%) for the SBD2 variants in absence and presence of 1 µM glutamine.

	L_{OA}, %	L_{OB}, %	L_C, %
SBD2	61.9 ± 0.3	59.3 ± 0.2	60.7 ± 0.2
SBD2 + Gln	61.7 ± 0.3	59.3 ± 0.3	60.5 ± 0.3
T256E	62.4 ± 0.2	59.4 ± 0.3	60.9 ± 0.4
T256E + Gln	62.4 ± 0.5	59.3 ± 0.6	60.8 ± 0.4
S358K	62.5 ± 0.5	59.9 ± 0.6	61.4 ± 0.7
S358K + Gln	62.6 ± 0.8	60.1 ± 0.9	61.9 ± 1.5
T256K	62.1 ± 0.7	59.6 ± 0.9	/
T256K + Gln	62.0 ± 0.9	59.6 ± 1.0	60.6 ± 1.0[A] 61.2 ± 1.1 [A]
S358E	61.0 ± 2.0	58.3 ± 2.0	59.5 ± 1.9
S358E + Gln	61.2 ± 2.7	58.6 ± 2.9	59.6 ± 2.8[A] 60.4 ± 2.8[A]
T256E+S358K	62.7 ± 0.7	60.1 ± 0.6	61.6 ± 0.7
T256E+S358K + Gln	63.4 ± 1.3	61.5 ± 0.3	62.2 ± 1.3
T256K+S358E	/	59.2 ± 0.2	61.3 ± 0.2
T256K+S358E + Gln	/	59.3 ± 0.4	60.9 ± 0.1

[A] Two different bound state levels were observed for SBD2(T256E) and SBD2(S358K) variants.

Table S3. Fitted kinetic parameters for SBD1

parameter	optimal fit values	95% confidence intervals	units
k_1	1.5	< 2.6	s^{-1}
k_{-1}	15	< 66 [A]	s^{-1}
k_2	14	> 0.4	$\mu M^{-1}\ s^{-1}$
k_{-2}	1.1×10^2	> 9.2	s^{-1}
k_{-2}/k_2 [B]	8.3	2.8 - 31	μM
k_3	1.4×10^2	> 58 [A]	s^{-1}
k_{-3}	6.5	4.3 – 10.0	s^{-1}

[A] It appears that the confidence intervals for k_{-1} and k_3 are overestimated because only a very small part of the experimental data points are influenced by these parameters, while the R^2 of the entire curve is analyzed to predict the intervals.

[B] It was noticed that the experimental data could be simulated with many values of k_2 and k_{-2}, providing their ratio was constant.

Table S4. Fitted kinetic parameters for SBD2.

parameter	optimal fit values	95% confidence intervals	units
k_1	3.6	< 7.3	s^{-1}
k_{-1}	56	29 - 98	s^{-1}
k_2	3.8	> 0	$\mu M^{-1}\ s^{-1}$
k_{-2}	2.2×10^2	> 18	s^{-1}
k_{-2}/k_2[A]	57	> 4.8	μM
k_3	2.5×10^3	> 4.3×10^2	s^{-1}
k_{-3}	37	31 - 40	s^{-1}

[A] It was noticed that the experimental data could be simulated with many values of k_2 and k_{-2}, providing their ratio was constant.

Table S5. Potential charge altering mutations as selected by bioinformatics.

original residue		frequency of occurrence (%) in natural homologs [a]		
in SBD2	position	aspartate	glutamate	lysine
M	254	14.8	7.4	
T	256		6.4	5.5
P	257		7.4	
K	258	11.8	16.7	
K	259	7.6	21.5	
D	260			54.2
V	261			8.8
T	263			18.0
E	274			
Q	276		10.7	8.0
N	277	38.0	11.3	
D	278			20.9
D	279			8.3
Q	281	10.0	36.0	30.7
T	283			12.0
N	291			32.7
A	292		9.3	
K	295		31.3	
N	296	44.0	18.7	
Q	297		20.7	
G	298	5.3		
K	300		19.3	
K	302	16.7	39.3	
N	304		6.0	50.7
G	307	15.3		
Q	309	34.0		
Q	316		27.3	
G	318	7.3		6.0

Table S5. Potential charge altering mutations as selected by bioinformatics.

original residue		frequency of occurrence (%) in natural homologs [a]		
in SBD2	position	aspartate	glutamate	lysine
A	332	9.3	74.0	
Q	335		12.0	37.3
V	336			26.0
S	341	60.7	15.3	5.3
S	345	27.3	30.7	12.7
S	353			31.3
S	354	5.3	20.7	45.3
T	355	43.9		
S	358	29.3	10.7	14.0
S	361			5.3
K	363	22.0	32.7	
K	366	6.7	10.7	
T	369			24.7
N	374			10.0
A	377		37.3	
F	379		6.7	
D	380			8.7
N	383		32.7	9.3
A	384	9.3	17.3	12.7
A	386		8.0	46.7
K	387	44.4	13.9	
E	388			64.6
G	390	6.8		
T	392		14.0	29.3
T	397	60.0	9.3	
T	400	24.0		
S	404	26.0	20.7	
S	405	38.0	32.0	

Table S5. Potential charge altering mutations as selected by bioinformatics.

original residue		frequency of occurrence (%) in natural homologs [a]		
in SBD2	position	aspartate	glutamate	lysine
N	407			38.7
N	412	42.7		
E	418			
G	428			5.3
K	430	10.7	7.3	
A	432		7.3	72.7
K	436	28.0	23.3	
P	437			32.7
I	438		49.3	
P	439	6.0	7.3	17.3
G	441	21.3	5.3	
Q	442		14.7	
S	451		17.8	17.8
P	453			14.0
E	454	11.3		
E	457	8.0		18.0
M	458			53.3
N	461	8.7	16.0	14.0
A	464			30.0
N	465	8.0	22.0	21.3
R	467			78.0
A	468	14.0	18.0	18.7
N	469	13.3		
E	471			18.7
D	473			5.3
K	474	5.3	40.7	
D	477			9.3
E	481			10.4

Table S5. Potential charge altering mutations as selected by bioinformatics.

original residue		frequency of occurrence (%) in natural homologs [a]		
in SBD2	position	aspartate	glutamate	lysine
S	482		41.8	12.7
D	483			7.3

[a] If the frequency that a D/E/K residue occurred at a given position was ≤ 5% it is not shown.

Table S6. Primers used for preparations of SBD2 variants.

SBD2 variant	Forward primer	Reverse primer
T256E	CAAGGTATGGCA***GAA***CCTAAAAA AGATG	GATTTAATCTGTATCAGG
S358K	ATGCCATAGCATTTTTATCC	CCATGACTTGAT***TTT***GTCATCT GTAC
T256K	CAAGGTATGGCA***AAA***CCTAAAAA AGATG	GGTTCGTCATCCATTAGAGC
S358E	ATGCCATAGCATTTTTATCC	CCATGACTTGAT***TTC***GTCATC TGTAC
T256E+S358K	CAAGGTATGGCA***GAA***CCTAAAAA AGATG	CCATGACTTGAT***TTT***GTCATCT GTAC
T256K+S358E	CAAGGTATGGCA***AAA***CCTAAAAA AGATG	CCATGACTTGAT***TTC***GTCATC TGTAC

References

(1) Flocco, M. M.; Mowbray, S. L. *J. Biol. Chem.* **1994**, *269* (12), 8931.

(2) Oswald, C.; Smits, S. H. J.; Höing, M.; Sohn-Bösser, L.; Dupont, L.; Le Rudulier, D.; Schmitt, L.; Bremer, E. *J. Biol. Chem.* **2008**, *283* (47), 32848.

(3) Berntsson, R. P. A.; Smits, S. H. J.; Schmitt, L.; Slotboom, D. J.; Poolman, B. *FEBS Lett.* **2010**, *584* (12), 2606.

(4) Shouldice, S. R.; Skene, R. J.; Dougan, D. R.; Snell, G.; McRee, D. E.; Schryvers, A. B.; Tari, L. W. *J. Bacteriol.* **2004**, *186* (12), 3903.

(5) Duan, X.; Quiocho, F. A. *Biochemistry* **2002**, *41* (3), 706.

(6) Huang, C. Y. *Methods Enzymol.* **1979**, *63* (C), 54.

(7) Kemmer, G.; Keller, S. *Nat. Protoc.* **2010**, *5* (2), 267.

(8) Howland, J. *Biochem. Mol. Biol. Educ.* **2001**, *29* (1), 36.

(9) Warwicker, J.; Charonis, S.; Curtis, R. A. *Mol. Pharm.* **2014**, *11* (1), 294.

(10) Bommarius, A. S.; Paye, M. F. *Chem. Soc. Rev.* **2013**, *42* (15), 6534.

(11) Biegert, a; Söding, J. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (10), 3770.

(12) Edgar, R. C. *BMC Bioinformatics* **2004**, *5*, 113.

(13) Edgar, R. C. *Nucleic Acids Res.* **2004**, *32* (5), 1792.

(14) Celniker, G.; Nimrod, G.; Ashkenazy, H.; Glaser, F.; Martz, E.; Mayrose, I.; Pupko, T.; Ben-Tal, N. *Isr. J. Chem.* **2013**, *53* (3–4), 199.

(15) Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. *Nucleic Acids Res.* **2010**, *38* (SUPPL. 2).

(16) Der, B. S.; Kluwe, C.; Miklos, A. E.; Jacak, R.; Lyskov, S.; Gray, J. J.; Georgiou, G.; Ellington, A. D.; Kuhlman, B. *PLoS One* **2013**, *8* (5).

(17) Vihinen, M. *Protein Eng.* **1987**, *1* (6), 477.

(18) Krieger, E.; Nielsen, J. E.; Spronk, C. A. E. M.; Vriend, G. *J. Mol. Graph. Model.* **2006**, *25* (4), 481.

(19) Krieger, E.; Darden, T.; Nabuurs, S. B.; Finkelstein, A.; Vriend, G. *Proteins Struct. Funct. Genet.* **2004**, *57* (4), 678.

(20) Wijma, H. J.; Floor, R. J.; Jekel, P. A.; Baker, D.; Marrink, S. J.; Janssen, D. B. *Protein Eng. Des. Sel.* **2014**, *27* (2), 49.

(21) Floor, R. J.; Wijma, H. J.; Colpa, D. I.; Ramos-Silva, A.; Jekel, P. A.; Szymański, W.; Feringa, B. L.; Marrink, S. J.; Janssen, D. B. *ChemBioChem* **2014**, *15* (11), 1660.

(22) Caves, L. S.; Evanseck, J. D.; Karplus, M. *Protein Sci.* **1998**, *7* (3), 649.

(23) Genheden, S.; Ryde, U. *J. Comput. Chem.* **2011**, *32* (2), 187.

(24) Wijma, H. J.; Marrink, S. J.; Janssen, D. B. *J. Chem. Inf. Model.* **2014**, *54* (7), 2079.

(25) Guerois, R.; Nielsen, J. E.; Serrano, L. *J. Mol. Biol.* **2002**, *320* (2), 369.

(26) Leaver-Fay, A.; O'Meara, M. J.; Tyka, M.; Jacak, R.; Song, Y.; Kellogg, E. H.; Thompson, J.; Davis, I. W.; Pache, R. A.; Lyskov, S.; Gray, J. J.; Kortemme, T.; Richardson, J. S.; Havranek, J. J.; Snoeyink, J.; Baker, D.; Kuhlman, B. *Methods Enzymol.* **2013**, *523*, 109.

(27) Miyazaki, K. *Methods Enzymol.* **2011**, *498*, 399.

(28) Gouridis, G.; Schuurman-Wolters, G. K.; Ploetz, E.; Husada, F.; Vietrov, R.; de Boer, M.; Cordes, T.; Poolman, B. *Nat. Struct. Mol. Biol.* **2014**, *22* (1), 57.

(29) Kapust, R. B.; Tözsér, J.; Fox, J. D.; Anderson, D. E.; Cherry, S.; Copeland, T. D.; Waugh, D. S. *Protein Eng.* **2001**, *14* (12), 993.

(30) Soskine, M.; Biesemans, A.; De Maeyer, M.; Maglia, G. *J. Am. Chem. Soc.* **2013**, *135* (36), 13456.

(31) Soskine, M.; Biesemans, A.; Moeyaert, B.; Cheley, S.; Bayley, H.; Maglia, G. *Nano Lett.* **2012**, *12* (9), 4895.

(32) Maglia, G.; Heron, A. J.; Stoddart, D.; Japrung, D.; Bayley, H. *Methods Enzymol.* **2010**, *475* (C), 591.

(33) Bonnet, G.; Krichevsky, O.; Libchaber, A. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (15), 8602.

(34) Elson, E. L.; Magde, D. *Biopolymers* **1974**, *13*, 1.

(35) Magatti, D.; Ferri, F. *Appl. Opt.* **2001**, *40* (24), 4011.

(36) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14* (1), 33.