Article

# Deep mutational scan of a drug efflux pump reveals its structure-function landscape

Received: 1 October 2021

Accepted: 10 October 2022

Published online: 28 November 2022

Check for updates

Gianmarco Meier<sup>1,5</sup>, Sujani Thavarasah <sup>©</sup><sup>1,5</sup>, Kai Ehrenbolger<sup>2,3</sup>, Cedric A. J. Hutter <sup>©</sup><sup>1,4</sup>, Lea M. Hürlimann <sup>©</sup><sup>1,4</sup>, Jonas Barandun <sup>©</sup><sup>2,3</sup> & Markus A. Seeger <sup>©</sup><sup>1</sup>⊠

Drug efflux is a common resistance mechanism found in bacteria and cancer cells, but studies providing comprehensive functional insights are scarce. In this study, we performed deep mutational scanning (DMS) on the bacterial ABC transporter EfrCD to determine the drug efflux activity profile of more than 1,430 single variants. These systematic measurements revealed that the introduction of negative charges at different locations within the large substrate binding pocket results in strongly increased efflux activity toward positively charged ethidium, whereas additional aromatic residues did not display the same effect. Data analysis in the context of an inward-facing cryogenic electron microscopy structure of EfrCD uncovered a high-affinity binding site, which releases bound drugs through a peristaltic transport mechanism as the transporter transits to its outward-facing conformation. Finally, we identified substitutions resulting in rapid Hoechst influx without affecting the efflux activity for ethidium and daunorubicin. Hence, single mutations can convert EfrCD into a drug-specific ABC importer.

Multidrug efflux pumps have been studied for more than four decades<sup>1</sup>. Drug efflux is an active transport process requiring the energy of ATP in the case of ABC transporters or ion gradients in the case of secondary-active transporters<sup>2</sup>. A hallmark of drug efflux pumps is their ability to transport a broad range of typically aromatic and cationic compounds. Structural analyses of drug efflux pumps and their regulator proteins revealed that drug binding pockets harbor aromatic amino acids to establish  $\pi$ - $\pi$  interactions with conjugated drug compounds<sup>3-7</sup>. However, in other instances, hydrogen bonds or salt bridges mediated by polar or charged residues were found to be of primary importance for drug recognition<sup>8,9</sup>. In contrast to the fast progress at the structural front, functional analyses lag behind, because drug transport assays are laborious and, thus, have to be limited to certain amino acids used for substitutions, as is the case for alanine scanning<sup>10</sup>.

Recent developments in next-generation sequencing (NGS) have enabled DMS analyses. DMS allows studying the effect of hundreds of variants with regard to protein activity in a single experiment by combining protein libraries with a selection procedure and NGS as a selection read-out<sup>11</sup>. DMS has thus far been mainly used to study protein–protein interactions and catalytic activity profiles of enzymes<sup>12,13</sup>.

In this study, we applied DMS to interrogate the sequence–function profile of EfrCD, a heterodimeric type I ABC exporter<sup>14</sup> (type IV ABC transporter, according to newer nomenclature<sup>15</sup>) stemming from the pathogenic bacterium *Enterococcus faecalis*. EfrCD exhibits strong drug efflux activity toward aromatic and cationic drugs, such as ethidium, Hoechst 33342 (referred to as Hoechst) and daunorubicin<sup>16</sup>. EfrCD consists of two transmembrane domains (TMDs) responsible for drug recognition and translocation and two nucleotide binding domains (NBDs), which couple the energy of ATP binding and hydrolysis to conformational changes at the TMDs. EfrCD features asymmetric ATP binding sites. The consensus site is mainly encoded by the EfrD NBD and is competent for ATP hydrolysis. The degenerate site can only bind but not hydrolyze ATP and allosterically regulates the consensus site<sup>17,18</sup>. Although the NBDs of EfrCD have been studied extensively<sup>17</sup>,

<sup>1</sup>Institute of Medical Microbiology, University of Zurich, Zurich, Switzerland. <sup>2</sup>Laboratory for Molecular Infection Medicine Sweden (MIMS), Department of Molecular Biology, Umeå Centre for Microbial Research, Umeå University, Umeå, Sweden. <sup>3</sup>Science for Life Laboratory, Umeå University, Umeå, Sweden. <sup>4</sup>Present address: Linkster Therapeutics AG, Zurich, Switzerland. <sup>5</sup>These authors contributed equally: Gianmarco Meier, Sujani Thavarasah. Se-mail: m.seeger@imm.uzh.ch



**Fig. 1** | **Cryo-EM structure of EfrCD. a**, Cartoon representation of EfrCD (EfrC in silver and EfrD in light yellow) in complex with nanobody Nb\_EfrCD#1 (in turquoise). **b**, Cross-sections of EfrCD to visualize the drug binding cavity. Amino acids subjected to DMS are colored in red (65 in the cavity and 14 in the NBDs).

it remains elusive how the TMDs recognize and extrude a broad spectrum of drugs out of the cell.

## Results

# Cryogenic electron microscopy structure of EfrCD and DMS library design

To identify residues lining the transmembrane cavity, we determined the structure of EfrCD using single-particle cryogenic electron microscopy (cryo-EM). The structure of EfrCD was obtained in complex with a nanobody, Nb\_EfrCD#1, at an overall resolution of 4.3 Å (Fig. 1a, Supplementary Figs. 1 and 2 and Supplementary Table 1). EfrCD exhibits a classical type I ABC exporter fold<sup>14</sup> and adopts an inward-facing conformation with the cavity exposed to the cytoplasm and the NBDs in close proximity to each other. The nanobody binds to the extracellular part of EfrCD and reaches into a crevice formed between EfrC and EfrD (Extended Data Fig. 1a). Nb EfrCD#1 binds with high affinity ( $K_0 = 7 \text{ nM}$ ; Extended Data Fig. 1b) and inhibits ATPase activity of EfrCD by approximately 70% (Extended Data Fig. 1c). Based on the cryo-EM structure of EfrCD, we identified 65 residues lining the large inward-facing cavity formed by the TMDs (Fig. 1b). To study their functional role, we mutated them one by one to all other 19 amino acids and, thereby, created a library of 1,235 cavity variants. In addition, we mutated 14 highly conserved residues of the asymmetric NBDs of EfrCD, which have been previously studied at the functional level<sup>17</sup>. In summary, our deep mutational scan included a total of 79 residues corresponding to 1,501 single variants.

#### DMS pipeline overview

To perform DMS experiments, three boundary conditions need to be met<sup>11,19</sup>. First, an input mutational library has to be generated (Fig. 2a, Supplementary Fig. 3 and Methods). Second, a genotype–phenotype linkage is needed to exert selection pressure on each library member. In our case, expression of EfrCD in *Lactococcus lactis* provides resistance toward toxic dyes, resulting in varying growth rates depending on which EfrCD variant is expressed in the individual cell (Supplementary Fig. 4 and Methods). The genetic information for the respective EfrCD variant is encoded on a plasmid (Fig. 2a,b). Our expression construct adds a green fluorescent protein (GFP) C-terminally to EfrD (Fig. 2a)

to measure EfrCD production levels. Third, NGS is employed to count each variant in the input library as well as after the selective growth to calculate variant scores (Fig. 2c, Supplementary Fig. 5 and Methods).

#### $Sequence-function\,map\,of\,EfrCD\,determined\,for\,three\,drugs$

A deep mutational scan of EfrCD was performed in the presence of daunorubicin, Hoechst or ethidium for three biological replicates and gave rise to sequence-function heat maps (Fig. 3). To normalize the data with respect to wild-type EfrCD, the efrCD gene was marked with a single base substitution, resulting in a silent mutation. L. lactis cells containing the respective plasmid were added at a frequency of 5% before the DMS experiment. For most variants, variant scores and corresponding standard errors based on the three biological replicates were calculated by the program Enrich2 (ref.<sup>20</sup>). For 0.3–0.7% of variants (frequency varies for the three drugs), depletion was so pronounced that they were not detected by NGS after competition. Consequently, variant scores could not be calculated in these cases of particularly strong depletion (dark blue squares in Fig. 3). In total, 186 variant scores were excluded from analysis (gray squares in Fig. 3) owing to at least one of the following reasons (Methods): (1) To account for growth biases of the studied variants, the DMS experiment was performed in the absence of drugs (Extended Data Fig. 2). Thereby, we identified 16 variants that grew clearly faster or slower than wild-type EfrCD (Extended Data Figs. 2 and 3a). (2) For approximately 3% of the variants, our NGS data suffered from near-cognate reading errors (see also Methods). (3) Some variant scores exhibited unacceptably high standard errors (>0.8) (Extended Data Fig. 3b-d). In summary, our DMS measurements allowed us to determine accurate values for 4,311 variant scores (namely, 1,434 for daunorubicin, 1,439 for Hoechst and 1,438 for ethidium) that exhibit excellent pairwise squared Pearson correlation coefficients greater than 0.82 among biological replicates (Supplementary Fig. 5e).

#### Validation of the DMS approach

As expected, the great majority of the NBD variants grew slower than wild-type in the presence of drugs (Fig. 4a). In agreement with a previously published study on EfrCD performed with single variants<sup>17</sup>, substitutions of the Walker A lysines (K369<sup>EfrC</sup> and K389<sup>EfrD</sup>), the



**Fig. 2** | **DMS pipeline for drug efflux pump EfrCD. a**, DMS library preparation. Three replicates of the DMS library were generated and used in all experiments. **b**, Competitive growth of *L. lactis* expressing EfrCD variants in the presence of three different drugs at sub-minimal inhibitory concentration (sub-MIC). c, NGS of input and competed libraries to determine variant scores for every variant.

D-loop aspartates (D498<sup>Efrc</sup> and D518<sup>EfrD</sup>), the consensus site Walker B glutamate (E512<sup>EfrD</sup>) and the consensus site switch loop histidine (H543<sup>EfrD</sup>) resulted in particularly strong depletion. The A-loop tyrosines (Y338<sup>EfrC</sup> and Y359<sup>EfrD</sup>), which provide base stacking interactions with the adenine ring of ATP, tolerated aromatic substitutions (Fig. 3). Intriguingly, many substitutions at the degenerate site ABC signature motif (T489<sup>EfrD</sup>) were beneficial, as was also shown previously at the example of the T489G<sup>EfrD</sup> single variant<sup>17</sup>. To directly correlate variant scores to growth<sup>21</sup>, we determined growth rates in the presence of ethidium for 28 individual variants. This analysis revealed an overall good correlation of expected versus measured growth rates (Extended Data Fig. 4).

#### Chemical hallmarks of the drug binding cavity

Substitution of residues in the TMDs lining the drug binding cavity were generally better tolerated than in the highly conserved NBDs. Interestingly, approximately 8% of TMD substitutions exhibited better ethidium resistance compared to wild-type EfrCD (variant score >0.2), whereas the corresponding frequencies were 4% and 3% for daunorubicin and Hoechst, respectively (Fig. 4a). This indicates that the drug efflux activity of EfrCD can be further improved by single amino acid substitutions.

Strikingly, the introduction of negative charges into the TMDs had an overall positive impact on ethidium resistance (Figs. 3 and 4b). Conversely, introducing positive charges resulted in a high proportion of depleted variants. Although all three drugs are positively charged at physiological pH (Fig. 2b), the positive charge of ethidium cannot be removed by deprotonation due to a quaternary nitrogen within the conjugated ring system<sup>22</sup>. This might explain why, in particular for ethidium, electrostatic attraction via the introduction of negatively charged glutamates or aspartates strongly increased drug efflux activity for many variants, whereas electrostatic repulsion via positively charged residues was poorly tolerated. The residues whose substitutions to negative charges resulted in enrichment upon ethidium exposure primarily locate in the upper half of the inward-facing binding cavity (Extended Data Fig. 5). Intriguingly, it does not appear to be relevant where exactly the beneficial negative charge is introduced, suggesting that these newly introduced negatively charged residues unlikely contribute to an increased binding affinity within a defined drug binding pocket but, rather, facilitate drug translocation along the cavity wall.

Next to charge compensation, aromatic residues were previously shown to play a key role in establishing  $\pi$ - $\pi$  and cation- $\pi$  interactions with conjugated ring systems and positive charges<sup>4,7</sup>. Although the deep mutational scan revealed a number of residues where the introduction of an aromatic amino acid was found to be enriched in the presence of at least two of the drugs (Fig. 3), our data show that the median variant score of the aromatic substitutions is in the same range as for polar-neutral and non-polar substitutions (Fig. 4b). Hence, opposed to negatively charged residues, additional aromatic residues do not generally improve the capacity of EfrCD to efflux drugs. Wild-type EfrCD contains five aromatic residues in the substrate binding pocket (F173<sup>EfrC</sup>, F180<sup>EfrC</sup>, F235<sup>EfrC</sup>, Y88<sup>EfrD</sup> and F255<sup>EfrD</sup>), of which only two (F180<sup>EfrC</sup> and F235<sup>EfrC</sup>) are of functional importance, as they did not tolerate substitutions well, with the exception of other aromatic amino acids (Fig. 3).

**Fig. 3** | **Sequence-function map of EfrCD.** Enrich2 software was used to calculate variant scores of the deep mutational scan of EfrCD (EfrC, top panel, and EfrD, lower panel) performed in the presence of daunorubicin (left), Hoechst (middle) or ethidium (right). Enriched variants are colored in red and depleted variants in blue. Diagonal lines in each square correspond to the standard error for the respective variant score calculated based on three biological replicates and are scaled such that the entire diagonal corresponds to a standard error of 0.8. Squares with dots have the wild-type amino acid at that position. Dark blue

Six residues form a high-affinity drug binding site

To identify structure-function relationships, the mean variant scores determined over all substitutions and drugs were visualized in the context of the EfrCD structure (Fig. 4c). This analysis revealed a cluster of six positions (F180<sup>EfrC</sup>, N228<sup>EfrC</sup>, G232<sup>EfrC</sup>, F235<sup>EfrC</sup>, D136<sup>EfrD</sup> and N143<sup>EfrD</sup>), which are particularly sensitive to mutations. This region. which we call the 'depleted cluster', is located in the lower half of the substrate binding cavity and involves residues of transmembrane helices (TMHs) 4 and 5 of EfrC and TMH 3 of EfrD (Fig. 5a). TMH 4 and TMH 5 of EfrC constitute the two domain-swapped helices, which cross over to EfrD and interact with the NBD of EfrD via the coupling helix, thus allowing for allosteric coupling with the consensus nucleotide binding site. Furthermore, this region becomes partially inaccessible as EfrCD transits from its inward-facing to its outward-facing conformation (Extended Data Fig. 6a-c). To further investigate the depleted cluster, we generated the corresponding single alanine variants for these six positions by site-directed mutagenesis. In addition, we randomly picked 41 depleted cluster variants from our DMS library (Supplementary Table 2). In a first analysis, we assessed growth in the presence of daunorubicin and expression levels via GFP fluorescence measurements of all depleted cluster variants. This analysis revealed that, independently of the substituting amino acid, the great majority of variants confer decreased daunorubicin resistance while most of them maintained normal expression levels (Supplementary Table 2).

In a second step, we thoroughly analyzed six variants with regard to (1) drug resistance in growth assays, (2) drug efflux based on Hoechst and ethidium fluorescence and (3) ATPase activity measurements using detergent-purified and nanodisc-reconstituted EfrCD variants (Fig. 5). In agreement with the DMS results, the ability of these six variants to grow in the presence of any of the three drugs was decreased (Fig. 5b-d and Extended Data Fig. 7). Next, we analyzed whether the variants were able to mediate Hoechst and ethidium efflux when overexpressed in L. lactis. To this end, we monitored the uptake of these fluorescent dyes in intact L. lactis cells (Extended Data Fig. 8a). This assay format is highly sensitive in distinguishing residual activities of strongly impaired variants<sup>17</sup>. We observed variable degrees of activity loss versus wild-type EfrCD, with two variants (G232AEfrC and F235Q<sup>EfrC</sup>) exhibiting almost completely abolished efflux activity akin to the ATPase-deficient E512Q<sup>EfrD</sup> control (Fig. 5e, f). Protein production levels based on GFP fluorescence revealed wild-type-like expression for five of the investigated variants, demonstrating that insufficient transporter production cannot explain their loss of function (Fig. 5g). D136A<sup>EfrD</sup> showed a 60% diminished production level relative to wild-type EfrCD but, at the same time, still conferred considerable resistance toward the three drugs (Fig. 5b-d), showing that a lower production level per se does not necessarily lead to a strong loss of function. To assess if the substitutions interfere with ATP hydrolysis, the respective variants were overexpressed in L. lactis and purified in detergent. In agreement with the GFP fluorescence data, all six variants could be purified and showed similar elution profiles as wild-type EfrCD when analyzed by size-exclusion chromatography (SEC) (Supplementary Fig. 6). The fraction of the main SEC peak eluting at around 11.4 ml and corresponding to heterodimeric EfrCD was used for reconstitution into nanodiscs, and basal ATPase activities were determined (Fig. 5h). With the exception of variant N143S<sup>EfrD</sup>, whose ATPase activity was diminished by approximately 40% relative

squares represent variants for which no NGS count was detected in at least one of the three competed libraries due to strong depletion. Gray squares denote variants for which no score was calculated. Randomized residues are indicated on the left of the heat map. NBD residues of the consensus nucleotide binding site are shaded in dark green and residues of the degenerate site in light green. The domain organization is depicted on the right. Substituting amino acids are indicated on the top and are grouped into positively charged (+), negatively charged (–), polar-neutral (P), non-polar (NP), aromatic (A) and unique (U).

to wild-type EfrCD, the ATPase activities of the remaining five depleted cluster variants were found to be intact, yet they exhibit decreased efflux activity. Our analyses, thus, strongly suggest that the residues

of the depleted cluster directly interact with the transported drugs and, thereby, define a high-affinity binding site for drugs when EfrCD adopts its inward-facing conformation.



#### Variants causing EfrCD-mediated Hoechst influx

In a search for specificity-determining residues, we screened the DMS data for EfrCD variants exhibiting a low variant score (less than -3) for one drug and a high variant score (greater than -0.2) for at least one other drug (Fig. 6a-c). This analysis gave rise to a cluster of residues whose substitution to mostly negatively charged residues resulted in Hoechst sensitivity (and, for I286D<sup>EfrC</sup>, in addition, daunorubicin sensitivity) while maintaining ethidium resistance similar to or better than wild-type EfrCD (Fig. 6b,c). Within this Hoechst-sensitivity cluster (Fig. 6d), we decided to analyze variants I239D<sup>EfrC</sup>, M243D<sup>EfrC</sup>, I286D<sup>EfrC</sup>, A32D<sup>EfrD</sup> and Q307D/E<sup>EfrD</sup> in detail (Extended Data Figs. 9 and 10a,b). In the in vivo fluorescence transport assay, these variants displayed an unexpectedly steep fluorescence increase upon Hoechst addition, which was clearly faster than Hoechst uptake in L. lactis cells expressing the ATPase-deficient E512QEfrD control variant (Fig. 6e and Extended Data Fig. 10a). This effect was Hoechst-specific, because the same set of variants behaved very similarly to wild-type EfrCD when monitoring ethidium uptake (Fig. 6f and Extended Data Fig. 10a). Interestingly, the variants did not accumulate the same amount of Hoechst as E512QEfrD. Instead, they reached a steady-state level, akin to the curves obtained with cells expressing wild-type EfrCD but at a higher intracellular Hoechst concentration. This suggests EfrCD-mediated influx of Hoechst along its concentration gradient at the beginning of the uptake experiment when the intracellular drug concentration is low. Once the intracellular concentration reaches a certain level, influx and efflux are in equilibrium with no net transport across the membrane. We propose that, in these variants, a secondary high-affinity binding site for Hoechst is created, which is exposed when the transporter adopts the outward-facing conformation (Extended Data Fig. 6d-f). The fact that the most pronounced Hoechst-sensitivity variants encompass substitutions with negatively charged residues further lends support to this hypothesis. Of note, similar observations were made previously on the major facilitator superfamily (MFS) multidrug efflux pump LmrP, where the introduction of cysteines at some positions resulted in a de-coupled transporter mediating facilitated ethidium influx into L. lactis cells<sup>23</sup>.

We next asked whether ATP-mediated conformational cycling at the NBDs is required for the fast uptake of Hoechst in these variants. To this end, we combined the Hoechst-sensitivity variant M243D<sup>EfrC</sup> with the ATPase-deficient E512Q<sup>EfrD</sup> variant. The M243D<sup>EfrC</sup>\_E512Q<sup>EfrD</sup> double variant exhibited an intermediate phenotype as compared to the respective M243D<sup>EfrC</sup> and E512Q<sup>EfrD</sup> single variants via a mechanism that requires further investigation (Fig. 6e).

#### Biochemical characterization of drug binding clusters

We sought to investigate the interplay between the two non-overlapping clusters identified based on DMS. To this end, we combined the Hoechst-sensitivity cluster variant M243D<sup>Efrc</sup> with depleted cluster variants F235Q<sup>Efrc</sup> or N143S<sup>EfrD</sup>. In the Hoechst uptake assay in intact cells, these double variants retained the steep initial increase in fluorescence

**Fig. 4** | **DMS analysis of EfrCD. a**, Percentage of enriched, neutral, depleted or strongly depleted variants for each domain (EfrC TMD, EfrC NBD, EfrD TMD and EfrD NBD) as determined by DMS in the presence of daunorubicin, Hoechst or ethidium. **b**, Box plot representation of variant scores (TMD residues only) for daunorubicin (blue), Hoechst (red) and ethidium (green) grouped by properties (aromatic (n = 190 variants), positively charged (+) (n = 177 variants), negatively charged (-) (n = 125 variants), non-polar (n = 244 variants), polar-neutral (n = 386 variants) and unique (n = 126 variants)) of the substituting amino acid. Shown is the median (colored line), the interquartile range (boxes) and the last data point in the 1.5-fold interquartile range (whiskers). The score of strongly depleted variants was arbitrarily set to -7. **c**, Mean variant scores of mutated residues across all three drugs are depicted as colored sticks on the structure of EfrC (left panel in silver) and EfrD (right panel in light yellow). EfrD is turned by 180° in relation to EfrC. The colors of mutated residues correspond to the scale shown on the right.

as it was observed for the M243D<sup>EfrC</sup> single variant (Fig. 6e), indicating rapid, facilitated Hoechst influx at the onset of the experiment. However, the steady-state Hoechst level when influx and efflux are in equilibrium was higher for both double variants than the M243D<sup>EfrC</sup> single variant. This is explained by the diminished capacity of the depleted cluster single variants F235Q<sup>EfrC</sup> and N143S<sup>EfrD</sup> to actively efflux Hoechst (Fig. 5e).





Fig. 5 | Single clone analysis of depleted cluster variants. a, Depleted cluster residues (colored in blue) localize on TMHs 4 and 5 of EfrC (silver) and on TMH 3 of EfrD (light yellow). b–d, Growth curves of depleted cluster variants determined in the presence of 8  $\mu$ M daunorubicin (b), 16  $\mu$ M ethidium (c) or 1.5  $\mu$ M Hoechst (d). e, f, Transport assay with fluorescent substrates Hoechst (e) and ethidium (f). Fluorescence spectroscopy was used to monitor the accumulation of drugs by *L. lactis* expressing the respective EfrCD variant. Drug efflux manifests in slower increase of fluorescence. g, Protein production level of depleted cluster variants relative to wild-type EfrCD (100%) as determined

Next, we conducted a Hoechst accumulation assay in inside-out vesicles (ISOVs) derived from *L. lactis* cells overexpressing the respective EfrCD variants. In this assay, the ISOVs are first incubated with Hoechst until a stable fluorescent signal is reached due to intercalation of Hoechst into the lipid bilayer, followed by the addition of ATP-Mg, which energizes EfrCD transporters that are facing with their NBDs to the outside. In case of active drug efflux, Hoechst accumulates in the vesicle lumen, which, in turn, leads to a decrease of fluorescent signal due to the concomitant acidification of the vesicle interior through the proton-translocating  $F_1F_0$ -ATPase and the protonation of Hoechst

by GFP fluorescence in *L. lactis* expressing the variants fused to GFP. **h**, Basal ATPase activity of purified EfrCD variants measured in nanodiscs. The signal from the inactive E512Q<sup>EfrD</sup> variant was subtracted, and shown is the ATPase activity relative to wild-type EfrCD (100%). Growth curves shown in **b**-**d** are representative data of biological triplicates (all curves shown in Extended Data Fig. 7). Fluorescence measurements shown in **e** and **f** are representative data of biological duplicates. Data in **g** and **h** correspond to mean ± standard deviations of technical triplicates. wt<sup>EfrCD</sup>, wild-type.

inside the vesicle lumen (Extended Data Fig. 8b)<sup>24</sup>. Hoechst transport mediated by the M243D<sup>Efrc</sup> variant was only mildly impaired relative to ISOVs containing wild-type EfrCD (Fig. 6g). This finding agrees with the notion that the M243D<sup>Efrc</sup> variant confers facilitated Hoechst influx along its concentration gradient (Fig. 6e) but is still capable of active Hoechst efflux, the latter being demonstrated by pumping Hoechst into the ISOVs (Fig. 6g). In contrast, the depleted cluster variant F235Q<sup>Efrc</sup> was almost completely inactive in terms of Hoechst efflux when being produced alone (Figs. 5e and 6g) or in the context of the M243D<sup>Efrc</sup>\_F235Q<sup>Efrc</sup> double variant (Fig. 6e,g). The depleted





ISOVs (**g**) containing overexpressed Hoechst-sensitivity variant M243D<sup>Efrc</sup> alone or in combination with depleted cluster variants F235Q<sup>Efrc</sup> or N143S<sup>EfrD</sup>. Shown are representative data of two biological replicates. **h**–**j**, Basal ATPase activity (**h**) and drug-modulated ATPase activity (**i**,**j**) of nanodisc-reconstituted EfrCD variants analyzed in **e**–**g**. ATPase activities were measured in the presence of increasing concentrations of Hoechst (**i**) or daunorubicin (**j**). Data were normalized to the basal activity of the respective variant in the absence of drugs. Data in **h**–**j** correspond to mean ± standard deviations of technical triplicates. **k**, Schematic data interpretation as explained in the main text. AU, arbitrary units; wt, wild-type. cluster variant N143S<sup>EfrD</sup> exhibited partial transport activity (Figs. 5e and 6g), which was further improved in the context of the respective M243D<sup>EfrC</sup>\_N143S<sup>EfrD</sup> double variant (Fig. 6e,g). This indicates allosteric cross-talk between the two clusters, in which the introduction of the M243D<sup>EfrC</sup> mutation in the Hoechst-sensitivity cluster improves the efflux activity of the N143S<sup>EfrD</sup> variant through a mechanism that requires further examination.

We purified these variants and reconstituted them into lipid nanodiscs to investigate the drug-induced modulation of ATPase activity, which is an indicator of specific drug-transporter interactions<sup>25,26</sup>. When ATPase activity is plotted against drug concentration, often a bell-shaped curve is observed. The prevailing explanation for this phenomenon is that the initial stimulation of ATPase activity at lower drug concentrations is due to the interaction of the substrate with a high-affinity binding site, whereas the inhibition of ATPase activity at higher substrate concentrations is due to the existence of secondary lower-affinity substrate binding sites within the transporter<sup>27,28</sup>. Nanodisc-reconstituted EfrCD variants exhibited basal ATPase activities similar to wild-type EfrCD (Fig. 6h). In agreement with previous work<sup>16</sup>, wild-type EfrCD displayed bell-shaped ATPase modulation curves for Hoechst and daunorubicin (Fig. 6i, j). The depleted cluster variants F235QEfrC and N143SEfrD exhibited weaker stimulation signals for Hoechst (Fig. 6i) or needed a higher drug concentration to reach maximal stimulation for daunorubicin (Fig. 6j), reinforcing our hypothesis that these variants exhibit diminished affinity at the polyspecific drug binding site (that is, the depleted cluster). Also, the M243D<sup>Efrc</sup> variant displayed a much reduced ATPase stimulation in the presence of Hoechst and daunorubicin throughout the entire range of drug concentrations (Fig. 6i, j). However, contrary to the F235QEfrC and N143S<sup>EfrD</sup> variants, the M243D<sup>EfrC</sup> variant has its stimulation maximum at a Hoechst concentration much lower than in the case of wild-type EfrCD. This observation might be explained by ATPase inhibition via occupation of the newly introduced Hoechst binding site already at a lower Hoechst concentration as compared to wild-type EfrCD (Fig. 6i). Of note, the effect was even more pronounced for daunorubicin, where the ATPase activity of the M243D<sup>EfrC</sup> variant can no longer be activated, which might suggest that daunorubicin inhibits ATPase activity via the newly introduced Hoechst-sensitivity site (Fig. 6j). However, neither the variant score nor single clone analysis of the M243DEfrC variant would suggest daunorubicin influx. Therefore, a plausible alternative explanation for the decreased Hoechst-stimulated or daunorubicin-stimulated ATPase activity of the M243DEfrC variant might be altered conformational coupling reactions between the TMDs and the NBDs directly caused by the M243D mutation. Interestingly, for the double variants M243DEfrC\_F235QEfrC and M243DE <sup>frC</sup>\_N143S<sup>EfrD</sup>, the ATPase activities at higher Hoechst or daunorubicin concentrations (>5  $\mu$ M) were consistently higher than for the single M243D<sup>EfrC</sup> variant, indicating allosteric cross-talk between the two binding sites.

Our mechanistic insights into EfrCD are summarized in Fig. 6k. Wild-type EfrCD encompasses an inward-oriented high-affinity drug binding site (depleted cluster) whose integrity is required for the efficient efflux of daunorubicin, Hoechst and ethidium. Mutations in this drug binding site (depleted cluster variants) decrease drug binding affinity and result in (partial) transporter inactivation. Introduction of negatively charged residues in the Hoechst-sensitivity cluster creates an outward-oriented secondary Hoechst binding site, which is responsible for rapid, facilitated influx of Hoechst into the cell without affecting the ability to actively efflux Hoechst. In comparison to wild-type EfrCD, the concomitant influx and efflux equilibrate at an increased intracellular Hoechst concentration. Finally, when variants of the two clusters are combined, rapid Hoechst influx encounters a diminished Hoechst efflux activity; thus, influx and efflux reach an equilibrium at further elevated intracellular Hoechst concentrations.

#### Discussion

Mutational studies have a long history in the field of ABC transporters. Of particular mention are the systematic mutational analyses of ABCB1, a project that took more than a decade to study the effect of single cysteine mutations introduced at the TMDs<sup>29</sup>. To increase efficiency, more recent studies have employed random mutagenesis and functional screening. Yeast ABC transporter Pdr5 mutants with altered drug specificity were identified<sup>30</sup>; the *Yarrowia lipolytica* pheromone ABC transporter Ste6 was trained to export the initially poorly exported a-factor of Saccharomyces cerevisiae<sup>31</sup>; and the resistance nodulation division (RND) drug efflux pump AcrB was screened for mutants that resisted inhibition by an efflux pump inhibitor<sup>32</sup>. Moreover, the 11 tryptophan residues of murine ABCB1 were screened for functionally neutral substitutions using site-saturation mutagenesis to generate a tryptophan-free, fully functional ABCB1 variant for drug interaction studies<sup>33</sup>. In three of these studies<sup>30–32</sup>, the mutations were introduced by random mutagenesis (that is, were untargeted). Furthermore, the screens operated based on positive selection-that is, only functionally neutral or gain-of-function variants were selected and further analyzed. Nevertheless, these pioneering studies provided the ground for this work, as they demonstrated the evolutionary flexibility of drug efflux pumps and ABC transporters.

Unfortunately, the randomness of the above approaches prevented a systematic recording of sequence–function maps for these transporters. The DMS approach presented here overcomes these limitations as it permits the in-depth characterization of a relatively small, but near-complete, set of single-site substitutions. Thereby, novel insights into the sequence–function landscape of the TMDs in terms of drug recognition and drug specificity can be obtained. We surmise that many of the identified patterns, and especially the Hoechst-sensitivity cluster, would have escaped discovery if a systematic alanine scan or a random mutagenesis approach had been chosen to analyze EfrCD. The key advantage of DMS is the quantification of both gain-of-function and loss-of-function substitutions under exactly identical conditions in a single culture flask, thereby overcoming the inherent day-by-day variability of growth experiments, as is evident from the three biological replicates of growth curves shown in Extended Data Figs. 7 and 9.

Our comprehensive dataset revealed that electrostatic interactions or repulsions play a very important role in drug efflux of charged drugs as the ones pumped by EfrCD. This is prominently seen for ethidium efflux that strongly benefits from the introduction of negative charges into the upper half of the inward-facing drug binding cavity. Strikingly, the exact position of the introduced negative charge does not appear to matter, indicating that charge interactions mainly play a role in facilitating drug flux along the transport pathway and not as part of high-affinity drug binding sites. Intriguingly, functional screens and targeted engineering of the multidrug MFS transporter MdfA revealed a number of positions where the introduction of negatively charged residues enabled the transport of mono-cationic and di-cationic drugs in an otherwise inactive transporter<sup>34,35</sup>. Because MdfA is a protondrug antiporter, the functional role of the negative charge was mainly attributed to the creation of a novel proton translocation pathway, next to the possibility that new electrostatic interactions were established between cationic drugs and the transporter.

Considering that  $\pi$ - $\pi$  and cation- $\pi$  interactions mediated by aromatic residues have been reported to be of particular importance to confer polyspecificity to drug efflux pumps<sup>4,6,7,36</sup>, we were surprised to observe that the introduction of additional aromatic residues into the drug binding cavity only occasionally resulted in improved drug resistance and was, on average, not more beneficial than the introduction of polar-neutral or non-polar residues (Fig. 4b). Nevertheless, two aromatic residues present in wild-type EfrCD (F180<sup>Efrc</sup> and F235<sup>EfrC</sup>) were found to be part of EfrCD's polyspecific drug binding site. This situation is reminiscent of the well-studied multidrug efflux pump AcrB, which features a deep binding pocket composed of several functionally crucial phenylalanines<sup>37,38</sup>, as well as ABCB1, whose drug binding pocket is lined with a large number of aromatic residues<sup>6,39</sup>.

In a remarkable study on Pdr5, an alanine mutation in the H-loop of the consensus ATP binding site altered the drug specificity of this fungal multidrug transporter<sup>40</sup>. In our DMS data on EfrCD, however, we did not find any substitutions within the NBDs that would alter substrate specificity. Another extensively discussed topic in the field is the asymmetry of heterodimeric ABC transporters, in particular such ones with degenerate ATP binding sites<sup>18,41,42</sup>. As expected, we clearly noted asymmetries at the level of the nucleotide binding sites. In contrast, variant scores did not greatly differ between EfrC and EfrD at the level of the TMDs (Fig. 4a), indicating that residues of both EfrC and EfrD contribute to drug recognition and transport.

The DMS analysis revealed a polyspecific high-affinity drug binding site, which is fully accessible only when EfrCD adopts its inward-facing state but is deformed and more narrow in EfrCD's outward-facing conformation (Extended Data Fig. 6a-c). This suggests drug extrusion akin to a peristaltic pump, a mechanism that was previously suggested for the RND drug efflux pump AcrB (ref. <sup>38</sup>) and ABCB1 (ref.<sup>6</sup>). The systematic DMS approach also revealed a vulnerability of EfrCD-namely, a cluster of residues in the upper part of the cavity whose substitution into negatively charged residues results in transporter variants that confer downhill Hoechst influx but still confer resistance toward ethidium. This cluster likely binds Hoechst with high affinity when EfrCD adopts its outward-facing state and, thereby, counteracts Hoechst extrusion mediated by the peristaltic pumping via the inward-oriented drug binding site (Fig. 6k). Hence, we discovered a delicate interplay between influx and efflux that finally decides over the ultimate direction of transport. Intriguingly, type I ABC exporters, which mediate siderophore or solute import, have been described recently<sup>26,43</sup>. Furthermore, ABCB1 containing 14 alanine substitutions in TMHs 6 and 12 was found to exhibit active drug import but lost its capacity to efflux drugs44.

Our work provides insights into the plasticity of the large substrate binding cavity of ABC transporters and shows how even single substitutions can influence transport directionality. The DMS data presented here form a solid basis to investigate the interplay of drug binding sites in more detail, and our robust DMS protocol constitutes a solid framework to gain novel molecular insights into drug efflux pumps in particular and membrane transporters in general.

#### **Online content**

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-022-01205-1.

# References

- Theodoulou, F. L. & Kerr, I. D. ABC transporter research: going strong 40 years on. *Biochem. Soc. Trans.* 43, 1033–1040 (2015).
- 2. Du, D. et al. Multidrug efflux pumps: structure, function and regulation. *Nat. Rev. Microbiol.* **16**, 523–539 (2018).
- Schumacher, M. A., Miller, M. C. & Brennan, R. G. Structural mechanism of the simultaneous binding of two drugs to a multidrug-binding protein. *EMBO J.* 23, 2923–2930 (2004).
- 4. Eicher, T. et al. Transport of drugs by the multidrug transporter AcrB involves an access and a deep binding pocket that are separated by a switch-loop. *Proc. Natl Acad. Sci. USA* **109**, 5687–5692 (2012).
- Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T. & Yamaguchi, A. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* 443, 173–179 (2006).
- Alam, A., Kowal, J., Broude, E., Roninson, I. & Locher, K. P. Structural insight into substrate and inhibitor discrimination by human P-glycoprotein. *Science* 363, 753–775 (2019).

- Le, C. A., Harvey, D. S. & Aller, S. G. Structural definition of polyspecific compensatory ligand recognition by P-glycoprotein. *IUCrJ* 7, 663–672 (2020).
- 8. Debruycker, V. et al. An embedded lipid in the multidrug transporter LmrP suggests a mechanism for polyspecificity. *Nat. Struct. Mol. Biol.* **27**, 829–835 (2020).
- 9. Heng, J. et al. Substrate-bound structure of the *E. coli* multidrug resistance transporter MdfA. *Cell Res.* **25**, 1060–1073 (2015).
- Saini, P. et al. Alanine scanning of transmembrane helix 11 of Cdr1p ABC antifungal efflux pump of *Candida albicans*: identification of amino acid residues critical for drug efflux. *J. Antimicrob. Chemother.* 56, 77–86 (2005).
- 11. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
- Zinkus-Boltz, J., DeValk, C. & Dickinson, B. C. A phageassisted continuous selection approach for deep mutational scanning of protein–protein interactions. ACS Chem. Biol. 14, 2757–2767 (2019).
- Romero, P. A., Tran, T. M. & Abate, A. R. Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc. Natl Acad. Sci. USA* **112**, 7159–7164 (2015).
- Srikant, S. & Gaudet, R. Mechanics and pharmacology of substrate selection and transport by eukaryotic ABC exporters. *Nat. Struct. Mol. Biol.* 26, 792–801 (2019).
- 15. Thomas, C. et al. Structural and functional diversity calls for a new classification of ABC transporters. *FEBS Lett.* **594**, 3767–3775 (2020).
- 16. Hürlimann, L. M. et al. The heterodimeric ABC transporter EfrCD mediates multidrug efflux in *Enterococcus faecalis*. *Antimicrob*. *Agents Chemother*. **60**, 5400–5411 (2016).
- Hürlimann, L. M., Hohl, M. & Seeger, M. A. Split tasks of asymmetric nucleotide-binding sites in the heterodimeric ABC exporter EfrCD. *FEBS J.* 284, 1672–1687 (2017).
- Hohl, M. et al. Structural basis for allosteric cross-talk between the asymmetric nucleotide binding sites of a heterodimeric ABC exporter. *Proc. Natl Acad. Sci. USA* **111**, 11025–11030 (2014).
- 19. Fowler, D. M. et al. High-resolution mapping of protein sequencefunction relationships. *Nat. Methods* **7**, 741–746 (2010).
- 20. Rubin, A. F. et al. A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* **18**, 150 (2017).
- 21. Kowalsky, C. A. et al. High-resolution sequence-function mapping of full-length proteins. *PLoS ONE* **10**, e0118193 (2015).
- 22. Luedtke, N. W., Liu, Q. & Tor, Y. On the electronic structure of ethidium. *Chemistry* **11**, 495–508 (2005).
- Mazurkiewicz, P., Poelarends, G. J., Driessen, A. J. M. & Konings, W. N. Facilitated drug influx by an energy-uncoupled secondary multidrug transporter. J. Biol. Chem. 279, 103–108 (2004).
- 24. Swain, B. M. et al. Complexities of a protonatable substrate in measurements of Hoechst 33342 transport by multidrug transporter LmrP. *Sci. Rep.* **10**, 20026 (2020).
- 25. Ambudkar, S. V. et al. Partial-purification and reconstitution of the human multidrug-resistance pump—characterization of the drug-stimulatable ATP hydrolysis. *Proc. Natl Acad. Sci. USA* **89**, 8472–8476 (1992).
- 26. Arnold, F. M. et al. The ABC exporter IrtAB imports and reduces mycobacterial siderophores. *Nature* **580**, 413–441 (2020).
- Al-Shawi, M. K., Polar, M. K., Omote, H. & Figler, R. A. Transition state analysis of the coupling of drug transport to ATP hydrolysis by P-glycoprotein. J. Biol. Chem. 278, 52629–52640 (2003).
- 28. Hegedus, C. et al. Ins and outs of the ABCG2 multidrug transporter: an update on in vitro functional assays. *Adv. Drug Delivery Rev.* **61**, 47–56 (2009).
- 29. Loo, T. W. & Clarke, D. M. Mutational analysis of ABC proteins. Arch. Biochem. Biophys. **476**, 51–64 (2008).

- Tutulan-Cunita, A. C., Mikoshi, M., Mizunuma, M., Hirata, D. & Miyakawa, T. Mutational analysis of the yeast multidrug resistance ABC transporter Pdr5p with altered drug specificity. *Genes Cells* 10, 409–420 (2005).
- Srikant, S., Gaudet, R. & Murray, A. W. Selecting for altered substrate specificity reveals the evolutionary flexibility of ATP-binding cassette transporters. *Curr. Biol.* **30**, 1689–1702 (2020).
- Schuster, S. et al. Random mutagenesis of the multidrug transporter AcrB from *Escherichia coli* for identification of putative target residues of efflux pump inhibitors. *Antimicrob. Agents Chemother.* 58, 6870–6878 (2014).
- Swartz, D. J. et al. Replacing the eleven native tryptophans by directed evolution produces an active P-glycoprotein with site-specific, non-conservative substitutions. *Sci. Rep.* **10**, 3224 (2020).
- Adler, J. & Bibi, E. Promiscuity in the geometry of electrostatic interactions between the *Escherichia coli* multidrug resistance transporter MdfA and cationic substrates. *J. Biol. Chem.* 280, 2721–2729 (2005).
- Tirosh, O. et al. Manipulating the drug/proton antiport stoichiometry of the secondary multidrug transporter MdfA. Proc. Natl Acad. Sci. USA 109, 12473–12478 (2012).
- Brown, K., Li, W. & Kaur, P. Role of aromatic and negatively charged residues of DrrB in multisubstrate specificity conferred by the DrrAB system of *Streptomyces peucetius*. *Biochemistry* 56, 1921–1931 (2017).
- Sjuts, H. et al. Molecular basis for inhibition of AcrB multidrug efflux pump by novel and powerful pyranopyridine derivatives. *Proc. Natl Acad. Sci. USA* **113**, 3509–3514 (2016).

- Seeger, M. A. et al. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. Science **313**, 1295–1298 (2006).
- Aller, S. G. et al. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323, 1718–1722 (2009).
- 40. Ernst, R. et al. A mutation of the H-loop selectively affects rhodamine transport by the yeast multidrug ABC transporter Pdr5. *Proc. Natl Acad. Sci. USA* **105**, 5069–5074 (2008).
- 41. Stockner, T., Gradisch, R. & Schmitt, L. The role of the degenerate nucleotide binding site in type I ABC exporters. *FEBS Lett.* **594**, 3815–3838 (2020).
- 42. Mishra, S. et al. Conformational dynamics of the nucleotide binding domains and the power stroke of a heterodimeric ABC transporter. *eLife* **3**, e02740 (2014).
- 43. Rempel, S. et al. A mycobacterial ABC transporter mediates the uptake of hydrophilic compounds. *Nature* **580**, 409–412 (2020).
- 44. Sajid, A. et al. Reversing the direction of drug transport mediated by the human multidrug transporter P-glycoprotein. *Proc. Natl Acad. Sci. USA* **117**, 29609–29617 (2020).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

 $\circledast$  The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

# Methods

#### Cloning

Vector pREXdmsC3GH was generated on the basis of pREXC3GH (Addgene, 47077) by introducing inverted SapI recognition sites, which allow to excise the *efrCD* operon by restriction digest using Sapl. Using primerspREXC3GH SapI for (5'-tatataGCAGGAAGAGCATTAGAAGTTT TG TTT CAA GGT CCA CAA TTC) and pREXC3GH\_SapI\_rev (5'-tat ata ACT AGA AGA GCC CAT GGT GAG TGC CTC CTT ATA ATT TAT TTT GTA G), the vector backbone was PCR-amplified, and the resulting fragment was ligated with the ccdB kill-cassette excised from pINIT cat (Addgene, 46858) with Sapl. Vector pREXNH3CA used to clone efrCD in frame with a C-terminal Avi-tag was constructed from pREXNH3 (Addgene, 47079) by PCR amplification with 5' phosphorylated primers pREXNH3(newAvi 5'P) FW (5'-aga aga tcg agt ggc acg agT AAT AAC TAG AGA GCT CAA GCT TTC TTT GA) and pREXNH3(newAvi 5'P) RV (5'-gag ctt cga aga tat cgt tca gac cTG CAG AAG AGC TGA ACT AGT GG), which insert the Avi-tag in front of the stop codon in pREXNH3. The resulting PCR product was circularized using blunt-end ligation. Using fragment exchange (FX) cloning and SapI restriction digest<sup>45</sup>, wild-type or mutant efrCD was first sub-cloned into pREXdmsC3GH or pREXNH3CA and, from there, via vector backbone exchange (VBEx) using Sfil restriction digest into the nisin-inducible L. lactis backbone pERL to finally obtain L. lactis expression vectors pNZdmsC3GH or pNZNH3CA, respectively<sup>46</sup>. Single clone variants were introduced by QuikChange site-directed mutagenesis on the efrCD-containing pREXdmsC3GH plasmid using the primers listed in Supplementary Table 3 and were verified by Sanger sequencing.

#### **DMS library preparation**

The single-site saturation libraries were synthesized by Twist Bioscience using the most abundant L. lactis codons (G:GGT, E:GAA, V:GTT, A:GCT, R:AGA, S:AGT, K:AAA, N:AAT, M:ATG, I:ATT, T:ACA, W:TGG, Y:TAT, L:TTA, F:TTT, C:TGC, Q:CAA, H:CAT, P:CCA and D:GAT) and overhangs containing SapI sites. Libraries were amplified by PCR using Phusion HF polymerase (Thermo Fisher Scientific) and PAGE-purified primers DMS EfrCD for 2 (5'-ACC ATG GGC TCT TCT AGT GAC CTT ATT C) and DMS\_EfrCD\_rev\_2(5'-TTC TAA TGC TCT TCC TGC TTC AAA AAC AAA TTG ATT TT). A sub-library per randomized position was generated by FX cloning<sup>45</sup> of these PCR products into the vector pREXdmsC3GH (see above) using chemically competent MC1061 cells. After recovery, transformed cells were grown overnight in LB medium containing 0.5% glucose and 100 µg ml<sup>-1</sup> of ampicillin. Transformation efficiency was determined by plating 1:100 of the recovery product onto LB agar plates containing 120 µg ml<sup>-1</sup> of ampicillin. Plasmid sub-libraries in the pREXdmsC3GH vector were extracted (QIAprep Spin Miniprep Kit, Qiagen). DNA concentration was determined based on A<sub>260</sub> using a NanoDrop. The sub-libraries in pREXdmsC3GH vector were mixed in the determined molar ratios (see Considerations for DMS library generation) for randomized positions in the *efrC* and *efrD* half-transporter genes, and three replicates (three independent pipetting reactions) for each half-transporter were created. Using VBEx cloning, the mixed sub-libraries were then sub-cloned into the L. lactis expression vector pNZdmsC3GH<sup>46</sup>. To this end, 200 µl of freshly prepared electro-competent L.lactis NZ9000 ΔlmrCD ΔlmrA cells were transformed with 200 ng of VBEx product, recovered and then grown in liquid M17 culture containing 0.5% glucose and 5 µg ml<sup>-1</sup> of chloramphenicol. Plating of small aliquots after recovery on M17 agar plates supplemented with 0.5 M sucrose, 0.5% glucose and 5 µg ml<sup>-1</sup> of chloramphenicol revealed that at least 10<sup>6</sup> colony-forming units were obtained per replicate. Thereby, the DMS library size of 1,501 variants was oversampled by more than 600-fold to prevent diversity bottlenecks. After overnight growth, glycerol stocks of the three DMS library replicates were prepared as 1-ml aliquots and stored at -80 °C. To be able to normalize our DMS data to the performance of wild-type EfrCD, a silent mutation at amino acid positon P240<sup>EfrC</sup> (CCA to CCT) was introduced.

#### **Considerations for DMS library generation**

Under ideal conditions, a DMS input library is evenly distributed. In praxis, however, this is not the case, because every library is to some extent skewed. Consequently, the NGS reads for the 1,501 DMS variants of the non-selected input library differ considerably. By increasing the sequencing depth, underrepresented variants can still be read with sufficient read counts but at increased costs. To achieve an optimal NGS read-out, we made major efforts to render the DMS library as evenly distributed as possible. In the process of library generation, we encountered and solved the following problems: (1) technical obstacles in commonly used methods to generate DNA libraries; (2) uneven distribution of randomized amino acids; (3) diversity bottlenecks; and (4) DNA shuffling arising from PCR amplification. For library generation, our first attempt was to use NNK degenerate primers and overlap PCR to amplify the *efrCD* gene and, thereby, generate sub-libraries for every randomized position<sup>47</sup>. This approach worked well at the technical level. However, the use of NNK degenerate primers resulted in skewed, unevenly distributed libraries, because the 20 amino acids are encoded by the 32 NNK codons. Hence, some amino acids (leucine, serine and arginine) are encoded by three codons and are, thus, overrepresented, whereas the other amino acids are encoded by two or a single codon. Another problem with the NNK randomization approach is that near-cognate sequencing errors bleed into the reads of randomized codons rather frequently. Therefore, we decided to order the single-site saturation libraries from a commercial provider (Twist Bioscience). These libraries have two technical hallmarks: (1) there is only one pre-defined codon for each amino acid (we used codons with highest abundance in L. lactis); and (2) the codons are guaranteed to be fairly evenly distributed for each randomized position (Supplementary Fig. 3d). For each of the 79 mutation sites in the efrCD gene, we obtained a separate sub-library from Twist Bioscience, which was sub-cloned one by one into the Escherichia coli pREXdmsC3GH cloning vector needed as intermediate step for VBEx cloning into a nisin-inducible L. lactis expression vector (Fig. 2a)<sup>46</sup>. In this cloning step, at least 10<sup>3</sup> colonies were obtained, from which the plasmids were prepared, to maintain the even distribution of the 19 variant codons for each of the mutation sites. Next, the DMS library was cloned from the pREXdmsC3GH vector into the L. lactis expression vector pNZdmsC3GH (Fig. 2a). The first open reading frame of efrCD variants cloned into this vector encodes for (untagged) EfrC, and the second open reading frame encodes for EfrD, followed by 3C protease cleavage site, GFP and a His<sub>10</sub>-tag. The GFP fusion allowed us to monitor expression levels in cells, whereas the His<sub>10</sub>-tag and the 3C cleavage site facilitate the purification of individual clones for biochemical characterization. To perform the cloning reaction, pREXdmsC3GH vectors containing the 79 mutation sites were first pooled and then sub-cloned. In a first attempt, we mixed the 79 pREXdmsC3GH sub-libraries at equimolar ratios. However, based on NGS, we realized that the 79 sub-libraries contained variable amounts of remaining wild-type efrCD genes, which most likely originate from library generation by Twist Bioscience using overlap PCR. This resulted in variable frequencies of variant codons for each of the 79 randomized positions (Supplementary Fig. 3b). To render the library more even in terms of variant codons in each of the 79 randomized positions, we took the NGS reads into account to adjust the mixing ratios of our 79 pREXdmsC3GH sub-libraries. As expected, this resulted in a more even distribution of variant codon frequencies of our library (Supplementary Fig. 3c). To measure replicates, we generated three separate DMS pools by mixing three times the respective pREXdmsC3GH vectors, followed by three independent VBEx cloning reactions to finally result in our DMS libraries in the L. lactis expression vector pNZdmsC3GH (which are referred to as the three biological replicates in this study) (Fig. 2a and Supplementary Fig. 3c). To avoid diversity bottlenecks and ensure an even distribution of the randomized codons, the number of colony-forming units was at least 500-fold greater than the DMS library size of 1,501 possible variants. Because VBEx is solely relying on

DNA digestion and ligation and avoids PCR amplification, DNA shuffling and consequential recombination of mutations were avoided<sup>48</sup>. Finally, individual cells might receive two (or more) plasmids during electroporation, and, consequently, such cells might express different variants at the same time. The L. lactis expression vector pNZdmsC3GH stems from the replicative multicopy L. lactis vector pNZ8048 (ref.<sup>49</sup>), meaning that multiple vectors encoding for different variants would not segregate as the cells divide. In an extreme case, a fully functional variant might mask the poor resistance phenotype of another variant. We estimated the rate of double transformations by electroporating two versions of pNZ8048 (one carrying a chloramphenicol resistance marker and the other one an erythromycin resistance marker<sup>50</sup>) into L. lactis under the same conditions as used for generating our DMS libraries. Cells were then plated on M17 agar plates containing chloramphenicol, erythromycin or both antibiotics, followed by colony counting, following the experimental approach described before<sup>21</sup>. In this way, we determined the double transformation rate to be 0.65% and 0.86% in the first and second technical repetition of the experiment, respectively. This number is likely to be overestimated, because we used intact plasmids instead of ligation product that was used to generate the library, and that results in a lower number of transformants per amount of DNA than circular plasmids. The problem of double transformation was further controlled by the generation of three completely independent input libraries (for which the double transformants would have different variant pairings) and which were used as biological triplicates to determine the final variant scores and standard errors. For the DMS experiments, we made three independent pre-cultures corresponding to these three independently prepared DMS libraries to inoculate medium containing nisin and the respective drug. Plasmids of these pre-cultures were isolated and sequenced by NGS to determine the variant counts of the input (or reference) library before selective growth.

#### Optimization of selective growth conditions to perform DMS

To achieve optimal competition conditions, we optimized the following parameters: (1) drug concentrations, (2) expression inducer concentration and (3) growth conditions. We used a test set of four EfrCD variants, including wild-type EfrCD and inactive E512Q EfrD variant as well as one variant with slightly increased efflux activity (F255AEfrD) and one variant with decreased efflux activity (I239AEfrC). Of note, these two alanine variants were identified when we performed a preliminary alanine scan of the upper region of the EfrCD drug binding cavity. First, the optimal nisin concentration used to induce efrCD expression was determined. A 1:10.000 (v/v) dilution of a nisin-containing L. lactis NZ9700 culture supernatant was found to be optimal, as it leads to high EfrCD overproduction (as measured based on fluorescence of GFP fused to the EfrD-chain) without affecting cellular growth (Supplementary Fig. 4a). Furthermore, we identified optimal drug concentrations by performing growth experiments with our test set of variants at various drug concentrations (Supplementary Fig. 4b). Based on this experiment, we decided to set the following optimal drug concentrations: daunorubicin (8  $\mu$ M), ethidium (16  $\mu$ M) and Hoechst (1.5  $\mu$ M). At these drug concentrations, we enrich for highly active variants (for example, F255A<sup>EfrD</sup>), partially deplete variants with somewhat decreased efflux activity (for example, I239AEfrc) and fully deplete inactive variants (for example, E512Q<sup>EfrD</sup>).

#### Competitive growth conditions used to perform DMS

For competition experiments, 50 ml of M17, 0.5% glucose with 5  $\mu$ g ml<sup>-1</sup> of chloramphenicol was inoculated with a glycerol stock (0.5-ml aliquot) of *L. lactis* containing the DMS library and grown overnight without shaking. Then, 50 ml of fresh M17 and 0.5% glucose with 5  $\mu$ g ml<sup>-1</sup> of chloramphenicol was inoculated with 500  $\mu$ l of the overnight cultures harboring the DMS libraries plus 10  $\mu$ l of overnight culture harboring wild-type *efrCD* containing a silent mutation at P240<sup>EfrC</sup>. Cells were grown for 2 hours at 30 °C without shaking. Protein expression was induced for 30 minutes by the addition of a nisin-containing culture supernatant of L. lactis NZ9700 added at a dilution of 1:10,000 (v/v).  $OD_{600}$  was determined and normalized to 0.5 in a volume of 2 ml. The remaining pre-culture was collected by centrifugation at 4.000g for 10 minutes and was used as the input library. Next, 24 ml of M17, 0.5% glucose, 5  $\mu$ g ml<sup>-1</sup> of chloramphenicol, containing nisin (1:10,000 (v/v)), with the respective drugs or without drug in case of the growth bias control experiment was inoculated with 1:100 induced L. lactis cells containing the DMS libraries ( $OD_{600}$  of 0.005 at the start of the competition). The following drug concentrations were used: 1.5 µM Hoechst 33342, 16 µM ethidium and 8 µM daunorubicin. Competitive growth was performed overnight at 30 °C without shaking. The next morning, an OD<sub>600</sub> of 2.5 was measured, meaning that, on average, the cells have divided approximately nine times. L. lactis cell pellets were collected by centrifugation at 4,000g for 10 minutes.

#### **Plasmid extraction**

The input and competed DMS libraries (encoded on the L. lactis plasmid pNZdmsC3GH) were extracted by resuspending the cell pellets in 200 µl of 20% sucrose, 10 mM Tris/HCl pH 8.0, 10 mM K-EDTA pH 8.0 and 50 mM NaCl, followed by the addition of 400 µl of 200 mM NaOH and 1% SDS. After inverting the tube four times, the solution was neutralized by the addition of 300 µl of 3 MK-acetate and 5 M acetic acid pH 4.8. After centrifugation at 20,000g for 10 minutes, 650 µl of supernatant was added to 1,350 µl of 100% EtOH and stored at -20 °C for 30 minutes. The precipitate was centrifuged at 4 °C and 20,000g for 5 minutes, and the pellet was washed with 500 µl of 70% EtOH. The pellet was dried at room temperature, resuspended in nuclease-free water containing 20 µg ml<sup>-1</sup> of RNaseA (Sigma-Aldrich) and incubated for 30 minutes at room temperature. The plasmids were purified using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). To extract the efrCD genes (encoding the DMS variants) for NGS, 7 µg of pNZdmsC3GH plasmid was digested with 40 U of Sfil (New England Biolabs) and separated on a 1% agarose gel, and the band containing the efrCD genes were extracted with the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel).

#### $Library\,preparation\,and\,NGS\,using\,Illumina\,NovaSeq\,6000$

In total, 500 ng of DNA was sheared twice in a Covaris microTUBE on a Covaris E220 Focused-ultrasonicator to achieve a distribution with an average fragment length of 150 base pairs (bp) using peak power 175, duty factor 10 and cycles per burst 200 for 280 seconds at 7 °C. In between cycles, the microTUBEs were spun down before second shearing. Fifty nanograms of sheared fragments was used as input for the TruSeq Nano DNA Library Prep Kit (Illumina). Illumina libraries were prepared following the manufacturer's protocol with one modification to yield a narrow fragment distribution optimized for 150-bp inserts by using a ratio of 5.4 parts of sample purification beads to one part of water (135  $\mu$ l of SPB + 25  $\mu$ l of nuclease-free water) in the library size selection step. Libraries were sequenced on a NovaSeq 6000 system (Illumina) using S1 flow cells and paired-end 2 × 151 cycles sequencing.

#### Data analysis

In a first step, adapter sequences, which are a result of the short insert size, were removed using Cutadapt (version 2.3). Adapter-trimmed reads were then aligned to the wild-type EfrCD DNA sequence using the Burrows–Wheeler Aligner (BWA) (version 0.7.17-r118). With a home-made Python program, the aligned sequences were then further filtered for sequences with a Phred quality score higher than 30. Reads with insertions or deletions were removed, and only overlapping sequences were allowed in the further processing. Variants were counted by allowing only reads with matching sequences in the overlapping region and with a single amino acid substitution. Variant scores were calculated using the software package Enrich2 (ref.<sup>20</sup>) (wild-type normalization and calculation of variant scores using natural log ratios).

#### NGS and data processing pipeline

A major technical challenge was the length of the *efrCD* genes (approximately 3.5 kb) and the fact that the mutated residues spread over the entire operon. This prevented us from using Illumina MiSeq (paired-end reads of approximately 250 bp), which is the predominant NGS method used in DMS analyses<sup>51-53</sup>. Instead, we decided to shear the *efrCD* gene into fragments of 130-180-bp length and to sequence them using the Illumina NovaSeq platform as paired-end reads with an overlap of 120-150 bp. To facilitate future DMS projects, we provide here detailed information on critical aspects of NGS library preparation, NGS and data analysis (Supplementary Fig. 5). The input material for NGS library construction are pNZdmsC3GH plasmids bearing the DMS library before and after competitive growth selection, which are extracted from L. lactis. In a first step, these plasmids are digested with Sfil to minimize the amount of backbone plasmid DNA in our NGS analysis. We noted that extraction of the transporter gene without any flanking sequences (using restriction enzymes cutting right at the beginning and end of the *efrCD* gene) resulted in very low sequencing coverage at the 5' region of efrC and the 3' region of efrD (Supplementary Fig. 3a). By using the Sfil digest, we include around  $2 \times 1,200$  bp of vector backbone upstream and downstream of the efrCD gene, which resolved this technical issue. Due to the large size of the efrCD transporter genes (3,482 bp), we then had to shear the DNA to be able to sequence them on Illumina sequencers. An option would be the PacBio technology, which, in theory, permits for full-length efrCD reads<sup>54</sup>. There are techniques that deal with sequencing of large gene variant libraries such as JigsawSeq; however, they rely on heavy randomization of the DNA sequence to generate overlaps to complete the 'jigsaw' and are, thus, not useful for our *efrCD* library with its variant positions being far apart from each other<sup>55</sup>. The latest generation of Illumina sequencers (such as the NovaSeq 6000 system) have Phred quality scores of Q30 in more than 90% of bases, corresponding to a base call accuracy of 99.9%. However, this accuracy was insufficient for our purposes, and we decided to rely on overlap paired-end sequencing results only (that is, we only took reads into consideration, which were read from both ends with the exact same sequence). For economic reasons, it was more attractive to use the Illumina NovaSeq 6000 system instead of Illumina MiSeq. To enable overlap paired-end sequencing, the sheared DNA fragments needed to have a size of approximately 150 bp, which is considerably shorter than in conventional Illumina library preparations (550 bp or 350 bp). With some adjustment of the standard Illumina library preparation protocol (see Library preparation and NGS using Illumina NovaSeq 6000), we achieved to obtain high-quality sequencing libraries with insert sizes of 150 bp. In our pipeline (Supplementary Fig. 5a), residual adapter sequences resulting from reading short fragments to the very end are removed. Reads are then aligned to the *efrCD* gene using the BWA. Aligned reads are quality-filtered by removing reads with low mean quality score (<30) and trimmed to the next codon start. Variants are called from the overlapping sequence of forward and reverse reads, thereby allowing only matching reads (that is, being identical in the forward and reverse read) and sequences with maximally one amino acid substitution with regard to the EfrCD wild-type sequence. The well-established DMS software Enrich2 is then used for wild-type normalization and calculation of variant scores using natural log ratios<sup>20</sup>. Despite considerable progress, sequencing errors of high-throughput sequencers still represent a challenge. Owing to the fact that the paired-end reads are more than 20 times shorter than the entire efrCD gene, the great majority of the sequenced fragments are identical to the efrCD wild-type sequence, and, with insufficient accuracy, sequencing errors would be spuriously attributed as mutations. The large number of 79 mutated positions caused an additional technical hurdle. Under idealized circumstances of a perfect DMS input library, NGS, on average, reads 78 times the wild-type efrCD sequence until a mutation is read, which then corresponds to one of the possible 19 substitutions. Hence, NGS reads need to be highly

redundant to achieve the required sequencing depth for a statistically significant read-out. In addition, sequencing errors (which cannot be completely suppressed even with overlap paired-end reading) result in near-cognate substitutions (exchange of a single base) of the frequently read wild-type codons, which can 'bleed' into the rarely read codons of true variants. To quantify the frequency and distribution of sequencing errors for our application, we performed an NGS analysis of the wild-type efrCD gene. To this end, we picked a single L. lactis colony propagating the *efrCD*-containing plasmid encoding for wild-type EfrCD, checked its correctness with Sanger sequencing and processed it with our DMS pipeline (two independent replicates, prepared and sequenced on two different days). This analysis revealed that only a small subset of our investigated variants (namely, 87 of 1,501) were affected by the near-cognate read bleeding issue (Supplementary Fig. 5b), whereas the remaining 1,414 variants exhibited false count rates of  $2 \times 10^{-5}$  or lower. When analyzing the location of the near-cognate read errors by plotting reading errors against the 79 randomized position of the DMS library on the efrCD gene, it becomes evident that this issue is highly site-specific (Supplementary Fig. 5d). Notably, the data of two independent NGS runs performed on different days with independently prepped wild-type efrCD genes are basically identical (Supplementary Fig. 5d). Finally, the read errors were analyzed according to all possible base substitutions (Supplementary Fig. 5c). In line with previous publications, we observed C > A and G > T conversions being the most frequently miscalled substitutions. When we then analyzed our DMS input libraries in light of the near-cognate read bleeding issue, we realized that, in approximately 3% of the investigated DMS variants, the sequencing errors accounted for more than 20% of the variant reads, which we considered as unacceptably high. Therefore, the corresponding enrichment scores were excluded, and the respective squares are colored in gray in the sequence-function maps (Fig. 3). Finally, we asked the question of how reliable our variant scores were determined. To analyze this, we calculated the variant scores independently for replicates 1, 2 and 3 (Fig. 2a) and plotted the values against each other (Supplementary Fig. 5e). The resulting pairwise squared Pearson correlation coefficients were in the ranges  $r^2 = 0.82 - 0.91$  for daunorubicin,  $r^2 = 0.86 - 0.94$  for Hoechst and  $r^2 = 0.93 - 0.94$  for ethidium (Supplementary Fig. 5e), which can be regarded as excellent $^{20}$ .

#### Detection and removal of variant score outliers

Ageneral problem of variant libraries is growth bias of variants even in the absence of the selective pressure, which then confound the results obtained under selective pressure. To account for this problem, we performed the DMS experiments in the presence of the inducer nisin but in the absence of drugs (Extended Data Fig. 2). When the resulting variant scores (corresponding to growth advantage/disadvantage relative to wild-type EfrCD) are plotted from the largest to the smallest value (Extended Data Fig. 3a), it becomes apparent that 16 residues show a particularly strong growth bias even in the absence of drugs. To take this bias into account, we excluded these 16 outliers from our downstream analysis (appearing as gray squares in Fig. 3), because their variant scores in the presence of drugs cannot be trusted (despite the fact that the corresponding values could be determined with good standard errors). As a further validation, we plotted all standard errors calculated by Enrich2 according to their value for each individual drug used (Extended Data Fig. 3b-d). This analysis revealed prominent outliers. We decided to remove 16 data points of low statistical quality. The respective variants are marked as gray squares in Fig. 3.

#### Analysis of single clone variants in growth assays

To determine growth curves of individual EfrCD variants, an overnight culture of *L. lactis* NZ9000  $\Delta lmrCD \Delta lmrA$  cells harboring mutants in the pNZdmsC3GH vector was used to inoculate 10 ml of M17, 0.5% glucose and 5 µg ml<sup>-1</sup> of chloramphenicol. Cells were grown for 2 hours at 30 °C, and then protein expression was induced by the addition of a

nisin-containing culture supernatant of *L. lactis* NZ9700 for 30 minutes (1:10,000 (v/v)). Cultures were normalized to  $OD_{600}$  of 0.5 (at a path length of 1 cm) and used to inoculate 150 µl of fresh medium (1:100 (v/v), M17, 0.5% glucose and 5 µg ml<sup>-1</sup> of chloramphenicol containing nisin (1:10,000 (v/v))) in 96-well plates.  $OD_{600}$  (at a path length of approximately 0.25 cm corresponding to the height of 150 µl of medium in a 96-well plate) was monitored every 10 minutes over at least 16 hours in a microplate reader at 30 °C with shaking between reads. Theoretical growth rates were calculated based on variant scores obtained by Enrich2. Variant scores were first transformed into enrichment ratios, and growth rates were calculated according to Kowalsky et al. (equation 9)<sup>21</sup>. Measured growth rates were determined in a time interval between 2 hours and 8 hours, using linear regression of Ln-transformed  $OD_{600}$  values.

#### Correlation of variant scores and growth rates

To validate our variant scores obtained via our DMS platform in a quantitative fashion, we determined the growth rate of 28 single variants relative to wild-type EfrCD in the presence of 16 µM ethidium as technical duplicates on two different days (biological replicates). Using the theoretical framework described in Kowalsky et al.<sup>21</sup>, we calculated expected growth rates based on the variant scores of these single variants determined by DMS (variant score of wild-type EfrCD is zero by definition). In these calculations, we assumed that the competitive DMS growth experiment was carried out for nine doublings, because nine doublings are required to expand the initial culture after inoculation ( $OD_{600} = 0.005$ ) to the culture having reached stationary phase ( $OD_{600}$  = 2.5). Calculated and measured growth rates as well as correlation plots for the two biological replicates are shown in Extended Data Fig. 4. Overall, the correlation was found to be good (squared Pearson correlation coefficients of 0.85 and 0.83 for the respective biological repetition). However, the variant scores obtained by DMS consistently overestimated (for enriched variants)/ underestimated (for depleted variants) the measured growth rates. The differences basically vanish if we assume a higher number of doublings (that is, 12) of the culture. The discrepancy of measured versus calculated growth rates can be explained by three possible scenarios (and a combination thereof). (1) It is plausible to assume that many cells died right away when they were exposed to the drugs, and the actual number of living cells at the onset of the competition experiment might have been much smaller than we estimated based on OD<sub>600</sub> measurements, and, accordingly, the number of generations to reach the final  $OD_{600}$  of 2.5 was likely higher. (2) The length of the lag-phase might depend on drug efflux activity of the variant, which would result in the overrepresentation of highly active variants and vice versa. (3) Cells expressing highly active variants might be less prone to die at the onset of the experiment when they encounter drugs in the competition experiment. Notably, the rank-order of variants in terms of growth rate is basically identical in the two assays (DMS experiment versus growth of individual variants). Therefore, we consider our DMS experiments as solidly validated.

# Determination of EfrCD production levels via GFP fluorescence

Induced cultures were prepared as described in the growth assay. Two milliliters of fresh M17, 0.5% glucose and 5  $\mu$ g ml<sup>-1</sup> of chloramphenicol containing nisin (1:10,000 (v/v)) was inoculated with induced cultures (1:100 (v/v)) and grown overnight at 30 °C. Cells were collected by centrifugation at 4,000*g* for 10 minutes and washed twice with PBS. Pellets were resuspended in PBS. OD<sub>600</sub> as well as GFP fluorescence (485-nm excitation and 528-nm emission) was measured in a microplate reader (Cytation 5 BioTek), and the fluorescence signal was normalized to OD<sub>600</sub>. After background subtraction of cells producing EfrCD without GFP tag (autofluorescence of *L. lactis*), expression levels relative to wild-type EfrCD were calculated.

To determine transport activity, *L. lactis* NZ9000  $\Delta lmrCD \Delta lmrA$  cells harbouring EfrCD variants in pNZdmsC3GH vector were grown overnight in M17 and 0.5% glucose containing 5 µg ml<sup>-1</sup> of chloramphenicol. Then, 40 ml of fresh medium was inoculated with overnight culture (1:50 (v/v)) and grown for 2 hours at 30 °C to OD<sub>600</sub> = 0.4–0.6. Protein expression was initialized by the addition of nisin-containing culture supernatant of *L. lactis* NZ9700 (1:1,000 (v/v)) and proceeded for 2 hours at 30 °C. Cells were collected by centrifugation (4,000*g*, 10 minutes, 4 °C), washed twice and resuspended in 4 ml of 50 mM K-phosphate pH 7.0 and 5 mM MgSO<sub>4</sub>. Transport assays were carried out exactly as described<sup>16</sup>.

#### **Purification of EfrCD variants**

EfrCD wild-type or variants were expressed in L. lactis NZ9000 AlmrCD ΔlmrA containing plasmid pNZNH3GS, pNZNH3CA or pNZdmsC3GH. Cells were grown in M17, 0.5% glucose and 5 µg ml<sup>-1</sup> of chloramphenicol at 30 °C to an OD<sub>600</sub> of 1, and expression was induced with a nisin-containing culture supernatant of L. lactis NZ9700 (1:5,000 (v/v)) for 4 hours. Membranes were prepared by passing cells four times through a microfluidizer at 35 kpsi in PBS pH 7.4 containing 15 mM K-EDTA pH 7.4. After low-spin centrifugation (8,000g, 10 minutes), 30 mM MgCl<sub>2</sub> was added to the supernatant, and the lysate was incubated with DNase for 30 minutes at 4 °C. After high-spin centrifugation (170,000g, 1 hour), the pellet was resuspended in 20 mM Tris-HCl pH 7.5 and 150 mM NaCl supplemented with 10% glycerol. Proteins were solubilized with 1% (w/v) n-dodecyl-β-D-maltoside (β-DDM) for 2 hours at 4 °C. Insolubilized fraction was removed by high-spin centrifugation (170,000g, 1 hour). The supernatant was supplemented with 30 mM imidazole pH 7.5 and loaded onto Ni-NTA columns with 2 ml of bed volume. When EfrCD was prepared for cryo-EM analyses, nanodisc reconstitutions and ATPase activity assays, detergent was exchanged to n-decyl- $\beta$ -D-maltoside ( $\beta$ -DM) by washing with 15 bed volumes of 50 mM imidazole pH 7.5, 200 mM NaCl, 10% glycerol and 0.3% (w/v) β-DM. Protein was eluted with 200 mM imidazole pH 7.5, 200 mM NaCl, 10% glycerol and 0.3% (w/v)  $\beta$ -DM and, for purifications from pNZNH3GS or pNZdmsC3GH, buffer was exchanged to 20 mM Tris-HCl pH 7.5, 150 mM NaCl with 0.3% (w/v) β-DM using PD10 columns, followed by overnight incubation with 3C protease. When EfrCD was prepared for alpaca immunizations for nanobody selections and analyses, the entire purification was carried out in 0.03% (w/v)  $\beta$ -DDM. For the biotinylated versions expressed from vector pNZNH3CA, the protein was concentrated to 360 µl using Amicon Ultra-4 concentrator units with 50-kDa molecular weight cutoff (MWCO). Biotinylation and 3C cleavage was performed overnight at 4 °C in a total reaction volume of 4 ml with 0.2 mg of 3C protease and 330 nM BirA (0.016 mg ml<sup>-1</sup>) in buffer containing 20 mM imidazole pH 7.5, 10 mM magnesium acetate, 200 mM ATP, 200 mM NaCl, 10% glycerol, 0.03% (w/v)  $\beta$ -DDM and a 1.2-fold molar excess of biotin.

For all constructs, cleaved His-tag and 3C protease were removed by reverse immobilized-metal affinity chromatography (IMAC), and EfrCD was polished by SEC on a Superdex 200 Increase 10/300 GL column in 20 mM Tris-HCl pH 7.5 and 150 mM NaCl, supplemented with 0.3% (w/v)  $\beta$ -DM or 0.03% (w/v)  $\beta$ -DDM.

#### Nanodisc reconstitution

Membrane scaffold protein MSP1E3D1 was expressed and purified as described <sup>56</sup>. Purified EfrCD proteins were reconstituted into nanodiscs at an EfrCD:MSP:lipid molar ratio of 1:14:500 in Na-HEPES pH 8.0 supplemented with 30 mM cholate. The reconstitution mixture was incubated at 4 °C for 30 minutes. Then, 200 mg of Bio-Beads SM-2 Resin (Bio-Rad) was added and incubated overnight at 4 °C while shaking at 650 r.p.m. After removal of Bio-Beads SM-2 Resin, the reconstituted transporter was further purified by SEC on a Superdex 200 10/300 GL column equilibrated with 20 mM Tris-HCl pH 7.5 and 150 mM NaCl.

#### ATPase activity assay

ATPase activity was measured using nanodisc-reconstituted EfrCD at a concentration of 2 nM in 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 10 mM MgSO<sub>4</sub>. For ATPase stimulation, Hoechst 33342 (at final concentrations ranging from 0.75  $\mu$ M to 48  $\mu$ M) or daunorubicin (at final concentrations ranging from 1.25  $\mu$ M to 80  $\mu$ M) were included. ATPase activities were measured at 30 °C for 15 minutes in the presence of 1 mM ATP, and liberated phosphate was detected colorimetrically using the molybdate/malachite green method. Then, 90  $\mu$ l of the reaction solution was mixed with 150  $\mu$ l of filtered malachite green solution (10.5 mg ml<sup>-1</sup> of ammonium molybdate, 0.5 M H<sub>2</sub>SO<sub>4</sub>, 0.34 mg ml<sup>-1</sup> of malachite green and 0.1% Triton X-100). Absorption was measured at 650 nm.

#### Nanobody selection and production

Nanobody Nb\_EfrCD#1 was generated by immunizing an alpaca with four subcutaneous injections of 200 µg of EfrCD in 20 mM Tris-HClpH 7.5, 150 mM NaCl and 0.03% (w/v) β-DDM in 2-week intervals. Immunizations of alpacas were approved by the Cantonal Veterinary Office in Zurich, Switzerland (animal experiment licence no. 188/2011). Phage libraries were generated as described previously<sup>57,58</sup>. Two rounds of phage display were performed on biotinylated EfrCD solubilized with 0.03% (w/v)  $\beta$ -DDM. After the last round of phage display, a 340-fold enrichment was determined by qPCR using AcrB as a negative control. The enriched library was then sub-cloned into the pSbinit vector by FX cloning, and 94 clones were screened with ELISA using biotinylated EfrCD as target. Of 83 positive hits, three distinct families based on the complementarity-determining region (CDR) composition were found. Nb EfrCD#1was cloned into expression plasmid pBXNPHM3 (Addgene, 110099) using FX cloning and expressed as described previously<sup>58</sup>. Tag-free Nb\_EfrCD#1 was used for cryo-EM structure determination.

#### Affinity determination by grating-coupled interferometry

The affinity of Nb\_EfrCD#1 was determined with grating-coupled interferometry (GCl) on the WAVEsystem (Creoptix AG). Avi-tagged EfrCD was captured on a streptavidin PCP-STA WAVEchip (polycarboxylate quasi-planar surface; Creoptix AG) to a density of 2,000 pg mm<sup>-2</sup>. For binding kinetics, Nb\_EfrCD#1 in 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.03% (w/v)  $\beta$ -DDM was injected with increasing concentrations (0.333, 1, 3, 9 and 27 nM) using 50  $\mu$ l min<sup>-1</sup> flow rate for 120 seconds at 25 °C, and dissociation was proceeded for 600 seconds. Data were analyzed on the WAVEcontrol (Creoptix AG) using double-referencing by subtracting the signals from blank injections and from the reference channel and fitted using a Langmuir 1:1 model.

#### **Cryo-EM structure determination**

The cryo-EM sample was prepared by purifying EfrCD, reconstituted in micelles, on a Superose 6 10/300 Increase GL size-exclusion column (GE Healthcare) in 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.15% (w/v)  $\beta$ -DM at 4 °C. Fractions from the monodisperse EfrCD peak were pooled, concentrated to 8 mg ml<sup>-1</sup> and supplemented with 5 mM MgCl<sub>2</sub>, 5 mM ATPyS and Nb\_EfrCD#1 in a 1:1 ratio to EfrCD. The supplemented sample was incubated for 20 minutes on ice before grid freezing. Quantifoil R2/1 Cu 200 grids were glow-discharged for 30 seconds at 15 mA before sample application and flash-freezing in a liquid ethane/propane mix using an FEI Vitrobot Mark IV (Thermo Fisher Scientific), set to a blot force of -5, waiting time of 1 second and blot time of 4.5 seconds at 100% humidity and 4 °C. Data were collected in three sessions on a Titan Krios (TFS) operated at 300 kV, equipped with a Gatan K2 BioQuantum direct electron detector at the Umeå Core Facility for Electron Microscopy. For automated data collection, EPU D-1.2 software (Thermo Fisher Scientific) was used to perform five acquisitions per hole using beam shifts. A total of 4,344 movies, each consisting of 40 frames, were collected with a dose of 52.1 e/Å<sup>2</sup>, a defocus range between  $-1.5 \,\mu$ m and  $-3.3 \,\mu$ m and a pixel size of 1.04 Å. Initial

movie alignment, drift correction and dose-weighting were done with MotionCor2. CTFFind-4.1.13 was used for contrast transfer function (CTF) determination. After manual inspection of the micrographs for drift or poor CTF fits, 3,135 micrographs remained. A total of 720,893 particles, picked using crYOLO, were extracted with a box size of 196 pixels (203.84 Å). A 30-Å lowpass-filtered initial model, obtained using cisTEM, was used as a reference in a three-dimensional (3D) classification with four classes using RELION-3.1. Particles from the top class (35.1%, 254, 682 particles) were re-extracted with a larger box size of 296 pixels (307.84 Å) and refined to a resolution of 5.50 Å. CTF refinement was performed in three subsequent steps. First, beam tilt, trefoil and 4th order aberrations were estimated, followed by the estimation of anisotropic magnification. Finally, defocus and astigmatism fitting was performed per micrograph. Next, the particles were subjected to Bayesian polishing using the trained sigma parameters of 1.263 Å per dose for velocity, 1,755 Å for divergence and 7.875 Å per dose for acceleration. The consensus 3D refinement resulted in a map of 4.73-Å resolution. The particle pool was exported into cryoSPARC version 3.2 and classified into three classes. Those classes were further subjected to heterogeneous refinement resulting in a top class with 212,083 particles. After non-uniform refinement, the final map reached an overall resolution of 4.25 Å (Supplementary Fig. 2). The resulting EM density map allowed for model building of the TMDs using the high-resolution X-ray structure of the ABC transporter TM287/288 (Protein Data Bank 4Q4H, 2.53 Å) as a template. The resolution in the peripheral regions (NBDs and nanobody) allowed for the placement of homology models. The resulting structure was optimized using ISOLDE and refined with PHENIX real\_space\_refine version 1.20.1-4487 against the final map at 4.25-Å resolution (Supplementary Table 1). The sidechain atoms in the less well-resolved peripheral regions (NBDs and nanobody) were removed from the final model. The directional Fourier shell correlation (FSC) determination was performed using 3DFSC, and model validation was performed according to ref.<sup>59</sup>. For model validation, all atomic coordinates were randomly displaced by 0.5 Å, followed by refinement against half map 1. The FSC coefficients of this refined model, and half map1 or half map2, were calculated using EMAN2. Model statistics are presented in Supplementary Table 1.

# Map and model visualization

Structure analysis and figure preparation were performed using Coot, PyMOL (Schrödinger) and UCSF ChimeraX (ref. <sup>60</sup>).

# **Reporting summary**

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

# Data availability

The cryo-EM map and coordinates for the EfrCD structure have been deposited in the EM Data Bank with accession code EMD-12816 and the Protein Data Bank with accession code 70CY. NGS datasets have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE189399. Source data are provided with this paper.

# **Code availability**

The code of the Python script used to analyze NGS data was deposited on https://github.com/giameier/DMS\_ABC.

#### References

- 45. Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry* **50**, 3272–3278 (2011).
- 46. Geertsma, E. R. & Poolman, B. High-throughput cloning and expression in recalcitrant bacteria. *Nat. Methods* **4**, 705–707 (2007).

#### Article

- 47. Kille, S. et al. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).
- Egloff, P. et al. Engineered peptide barcodes for in-depth analyses of binding protein libraries. Nat. Methods 16, 421–428 (2019).
- Kunji, E. R. S., Slotboom, D. J. & Poolman, B. Lactococcus lactis as host for overproduction of functional membrane proteins. *Biochim. Biophys. Acta* 1610, 97–108 (2003).
- 50. Guffick, C. et al. Drug-dependent inhibition of nucleotide hydrolysis in the heterodimeric ABC multidrug transporter PatAB from *Streptococcus pneumoniae*. *FEBS J.* **289**, 3770–3788 (2022).
- Starita, L. M. & Fields, S. Deep mutational scanning: library construction, functional selection, and high-throughput sequencing. *Cold Spring Harb. Protoc.* 2015, 777–780 (2015).
- Newberry, R. W., Leong, J. T., Chow, E. D., Kampmann, M. & DeGrado, W. F. Deep mutational scanning reveals the structural basis for α-synuclein activity. *Nat. Chem. Biol.* 16, 653–659 (2020).
- Jones, E. M. et al. Structural and functional characterization of G protein-coupled receptors with deep mutational scanning. *eLife* 9, e54895 (2020).
- 54. Rhoads, A. & Au, K. F. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics **13**, 278–289 (2015).
- Cho, N. et al. De novo assembly and next-generation sequencing to analyse full-length gene variants from codon-barcoded libraries. *Nat. Commun.* 6, 8351 (2015).
- 56. Hutter, C. A. J. et al. The extracellular gate shapes the energy profile of an ABC exporter. *Nat. Commun.* **10**, 2260 (2019).
- Zimmermann, I. et al. Synthetic single domain antibodies for the conformational trapping of membrane proteins. *eLife* 7, e34317 (2018).
- Zimmermann, I. et al. Generation of synthetic nanobodies against delicate proteins. *Nat. Protoc.* 15, 1707–1741 (2020).
- Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 136–153 (2015).
- 60. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

# Acknowledgements

We wish to thank all members of the Seeger and Barandun laboratories for scientific discussions. G.M. and M.A.S. thank the Functional Genomics Center Zurich (M. D. Moccia, L. Poveda and W. Qi) for their assistance with deep sequencing. We thank S. Štefanić of the Nanobody Service Facility, University of Zurich, for alpaca immunization. C. Perez is acknowledged for initial help with grid freezing of EfrCD. The electron microscopy data were collected at the Umeå Core Facility for Electron Microscopy, a node of the Cryo-EM Swedish National Facility, funded by the Knut and Alice Wallenberg Foundation, the Erling-Persson Family Foundation, the Kempe Foundation, SciLifeLab, Stockholm University and Umeå University. J.B. acknowledges funding from the Swedish Research Council (2019-02011), the SciLifeLab National Fellows program and Molecular Infection Medicine Sweden. This work was funded by a Swiss National Science Foundation Professorship (PP00P3\_144823, to M.A.S.), a Swiss National Science Foundation Project Grant (310030\_188817, to M.A.S.), a European Research Council (ERC) Consolidator Grant (MycoRailway, no. 772190, to M.A.S.) and an ERC Starting Grant (PolTube, no. 948655, to J.B.).

## **Author contributions**

G.M. and M.A.S. conceived the project. G.M. generated the DMS libraries, established the selection protocol, programmed and validated the NGS data analysis pipeline and performed the ATPase activity assays. S.T. generated the great majority of the single clone variants and analyzed them in growth assays and fluorescence transport assays, together with G.M. Nanobodies were generated by G.M. and L.H. and analyzed by G.M. and S.T. Cryo-EM analyses were performed by K.E., under the supervision of J.B., with EfrCD protein purified by C.A.J.H. Cryo-EM data analysis and model building were performed by K.E. and J.B. L.H. generated preliminary mutational data on alanine variants within the TMDs and supervised G.M. Figures were made and edited by G.M., S.T., K.E., J.B. and M.A.S. All authors wrote the paper.

# **Competing interests**

The authors declare no competing interests.

# **Additional information**

**Extended data** is available for this paper at https://doi.org/10.1038/s41589-022-01205-1.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41589-022-01205-1.

**Correspondence and requests for materials** should be addressed to Markus A. Seeger.

**Peer review information** *Nature Chemical Biology* thanks Parjit Kaur, Dirk Slotboom and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at www.nature.com/reprints.





**Extended Data Fig. 1** | **Characterization of nanobody Nb\_EfrCD#1. a**, Cartoon representation of Nb\_EfrCD#1 binding to the extracellular part of EfrCD. Complementary determining regions (CDRs) 1, 2 and 3 are colored in yellow, orange and red, respectively. **b**, Affinity determination of Nb\_EfrCD#1 using grating-coupled interferometry (GCI). EfrCD was immobilized on a WAVEchip

and Nb\_EfrCD#1 was injected at 0.33, 1, 3, 9 and 27 nM. A 1:1 kinetic binding model was used for data fitting (black curve) and values for on-rate ( $k_a$ ), off-rate ( $k_d$ ) and dissociation constant ( $K_D$ ) are given in the graph. **c**, ATPase activity of detergent-solubilized EfrCD in the presence and absence of 1  $\mu$ M Nb\_EfrCD#1. Shown is the mean and standard deviation of technical triplicates.



#### Variant score

**Extended Data Fig. 2** | **Variant scores determined in the absence of drugs.** Control experiment in which the DMS libraries were grown in the absence of drugs under otherwise identical conditions as in the competitive growth experiments shown in Fig. 3 (that is, in the presence of inducer nisin). Enrich2 was used to calculate variant scores based on the data obtained from three biological replicates. Enriched variants are colored in red and depleted variants in blue. The same scaling as in Fig. 3 was applied. Diagonal lines in each square correspond to the standard error for the variant score and are scaled such that the entire diagonal corresponds to a standard error of 0.8. Squares with dots have the wildtype amino acid at that position. Gray squares denote variants for which no score was calculated due to high errors or sequencing biases. Randomized residues are indicated on the left of the heat map. NBD residues of the consensus nucleotide binding site are shaded in dark green and residues of the degenerate site in light green. Substituting amino acids are indicated on the top and are grouped into positively charged (+), negatively charged (–), polar-neutral (P), non-polar (NP), aromatic (A) and unique (U).





**Extended Data Fig. 3** | **Variant filtering. a**, Variant scores determined in the absence of drugs (see Extended Data Fig. 2) were ordered according to the score value. Annotated variants colored in blue deviate strongly from wild-type EfrCD (variant score = 0) and were excluded in the DMS analysis (gray squares in Fig. 3).

**b**, **c**, **d**, Standard errors calculated by Enrich2 for all variants ordered according to their value for daunorubicin (**b**), Hoechst (**c**) and ethidium (**d**). Variants exhibiting suspiciously high standard errors are colored in blue and annotated. These variant scores were excluded in the DMS analysis (gray squares in Fig. 3).

а





b

С

Variant	Growth rate calculated based on variant score (rel. to wt <sup>ErcD</sup> )	Measured growth rate (rel. to wt <sup>EfrCD</sup> )	
variant		Replicate 1	Replicate 2
N124R <sup>Efrc</sup>	0.63	0.89	0.67
F180A <sup>EfrC</sup>	0.73	0.80	0.77
F180T <sup>EfrC</sup>	0.78	0.82	0.68
N228A <sup>EfrC</sup>	0.62	0.70	0.62
239A <sup>EfrC</sup>	0.76	0.80	0.80
239D <sup>EfrC</sup>	1.14	1.07	1.16
239L <sup>EfrC</sup>	0.93	1.02	0.90
M243D <sup>EfrC</sup>	1.16	1.17	1.19
286N <sup>EfrC</sup>	1.03	0.94	0.96
287P <sup>EfrC</sup>	0.95	0.95	0.92
R297A <sup>EfrC</sup>	0.59	0.78	0.61
A32D <sup>EfrD</sup>	1.10	1.10	1.11
D136A <sup>EfrD</sup>	0.37	0.63	0.60
S141E <sup>EfrD</sup>	0.69	0.52	0.75
N143A <sup>EfrD</sup>	0.46	0.52	0.45
N143D <sup>EfrD</sup>	1.27	1.04	0.98
N143S <sup>EfrD</sup>	0.47	0.68	0.55
	0.62	0.65	0.67
Q148D <sup>EfrD</sup>	1.24	1.12	1.00
Q148E <sup>EfrD</sup>	1.27	1.11	0.99
T151D <sup>EfrD</sup>	1.29	1.20	1.02
T151E <sup>EfrD</sup>	1.27	1.13	0.97
-255A <sup>EfrD</sup>	0.99	0.95	0.92
Q303E <sup>EfrD</sup>	1.25	1.08	0.99
M306D <sup>EfrD</sup>	1.24	1.12	1.00
Q307D <sup>EfrD</sup>	1.07	1.10	1.07
	1.31	1.21	1.05
	0.79	0.81	0.69

**Extended Data Fig. 4** | **Correlation analysis of calculated versus measured growth rates.** Growth rates for 28 individual variants relative to wild-type EfrCD were determined in the presence of ethidium as two biological replicates. Each growth rate data point is the average of at least two technical replicates. Theoretical growth rates were calculated based on variant scores (see Methods). **a**, **b**, Pearson correlation analysis of measured versus calculated growth rates for the two biological replicates, respectively. **c**, List of analyzed variants and the corresponding calculated and measured growth rates.



**Extended Data Fig. 5** | **Substitutions towards negatively charged residues enriched in the presence of ethidium.** Surface representation of EfrCD cut into two halves to visualize the substrate binding cavity. Residues with variant scores >1 in the presence of ethidium when substituted to aspartate or glutamate are depicted in red.



#### Hoechst-sensitivity cluster

Extended Data Fig. 6 | Depleted cluster and Hoechst-sensitivity cluster in the structural context of outward-facing EfrCD. Side-by-side representation of inward-facing EfrCD structure (a and d) and outward-facing EfrCD homology model (**b**, **c**, **e** and **f**) based on the coordinates of Sav1866 (PDB: 2HYD). The depleted cluster residues are colored in blue (**a**-**c**) and the Hoechst-sensitivity cluster residues in purple (**d**-**f**).





type EfrCD (wt<sup>EfrCD</sup>, dashed line) or the inactive E512Q<sup>EfrD</sup> variant (dotted line) were included as controls. The growth experiment was carried out on three separate days, resulting in three biological replicates.

#### а

Intact cells

b

Inside-out vesicles (ISOVs)





**Extended Data Fig. 8** | **Schematic drawing of Hoechst assays.** Hoechst accumulation assay in intact cells. Hoechst fluorescence increases due to intercalation of Hoechst into the chromosomal DNA and into the lipid bilayer. Active Hoechst efflux mediated by EfrCD results in a slower increase of fluorescence. b, Hoechst transport into inside-out vesicles (ISOVs). ATP is added

from the outside to pump Hoechst into the vesicle lumen. ATP is also consumed by the  $F_{\rm I}F_{\rm 0}$ -ATP ase to pump protons into the ISOV lumen, thereby acidifying the intraluminal milieu. Hoechst fluorescence decreases as a result of protonation of Hoechst due to the lower pH inside the ISOV.

#### https://doi.org/10.1038/s41589-022-01205-1





type EfrCD (wt<sup>EfrCD</sup>, dashed line) or the inactive E512Q<sup>EfrD</sup> variant (dotted line) were included as controls. The growth experiment was carried out on three separate days, resulting in three biological replicates.



а



**Extended Data Fig. 10** | **Single clone analysis of Hoechst-sensitivity variants. a**, Accumulation of fluorescent drugs Hoechst (upper row) or ethidium (lower row) in intact *L. lactis* NZ9000  $\Delta lmrCD \Delta lmrA$  expressing Hoechst-sensitivity variants. Wild-type EfrCD and inactive E512Q<sup>EfrD</sup> variant were included in all measurements. Per individual graph, the experiment performed on the same day is shown. All fluorescence experiments were carried out twice on the same

day (technical duplicates) and were performed on two separate days with freshly prepared cells (biological replicates). Representative results from these four replicates are shown. **b**, Expression levels of variants based on GFP fluorescence of *L. lactis* NZ9000 *ΔlmrCD ΔlmrA* expressing transporter variants containing GFP fusion tags. Data are represented as mean + /- standard deviations of technical triplicates.

# nature portfolio

Corresponding author(s): Markus A Seeger

Last updated by author(s): Oct 5, 2022

# **Reporting Summary**

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

# Statistics

For	all st	atistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.
n/a	Cor	firmed
	$\boxtimes$	The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
	$\boxtimes$	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
$\boxtimes$		The statistical test(s) used AND whether they are one- or two-sided Only common tests should be described solely by name; describe more complex techniques in the Methods section.
$\boxtimes$		A description of all covariates tested
$\boxtimes$		A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
	$\boxtimes$	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
$\boxtimes$		For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable</i> .
$\boxtimes$		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
$\boxtimes$		For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
	$\boxtimes$	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
		Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.

# Software and code

Policy information	about <u>availability of computer code</u>	
Data collection	NovaSeq Control Software v1.7.5, EPU-D 1.2(Thermo Fisher Scientific), WAVEcontrol 4.0(Creoptix AG)	
Data analysis	A custom Python script was used to analyse NGS data, which was deposited on https://github.com/giameier/DMS_ABC.	
	Standard programs for NGS analysis: Cutadapt 2.3, Burrows-Wheeler Aligner 0.7.17, Enrich2 v1.2.0	
	Standard programs for cryo-EM analysis: MotionCor2 2_1.2.6, CTFFind-4.1.13, crYOLO-1.3.5, cisTEM 1.0.0-beta, RELION-3.1, cryoSPARC v3.2, Coot-0.8.9.2-preEL (revision count 7607), PHENIX 1.14-3260, EMAN-2.22 final, 3DFSC (https://3dfsc.salk.edu), Chimera 1.13.1 (build 41965), ChimeraX 0.93, Pymol 1.8.2.3,	
	Program for GCI data analysis: WAVEcontrol 4.0 (Creoptix AG)	

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The cryo-EM map and coordinates for the EfrCD structure have been deposited in the EM Data Bank with accession code EMD-12816 and the Protein Data Bank with accession code PDB-7OCY. NGS data (raw and processed) have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Seriesaccession number GSE189399.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Ecological, evolutionary & environmental sciences

Behavioural & social sciences For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	DMS: A DMS library containing a total of 1501 variants was generated. Cloning reactions were performed with great redundancy (oversampling of at least 500 fold as determined by colony forming unit (cfu) counting) to avoid diversity bottlenecks. We took measures to render the input library as equal as possible, which ensured that all variants were detected by Next Generation Sequencing and that sufficient reads were determined for accurate variant scoring.
	Cryo-EM: The cryo-electron microscopy (cryo-EM) sample size was determined by the allocation of microscope time and particle density on the cryo-EM grids. The sample size was determined by the number of particles that were used for 3D classification (n=720893), which was sufficient to obtain a cryo-EM map with clear side chain features.
Data exclusions	DMS: For around 3% of variants, our NGS data suffered from near-cognate reading errors and the respective scores were not calculated. Further, 16 variants exhibited strong growth bias in the absence of drugs, and therefore were excluded in the DMS analyses. Finally, a total of 16 variant scores were removed from the DMS analysis owing to unreasonably large standard errors among the three biological replicates.
	Cryo-EM: Micrographs with poor CTF fits, ice of low quality, or drift, have been removed according to established standard procedures.
Replication	DMS: We independently generated the DMS library three times and used these three biological replicates for all experiments. High correlation between these three biological replicates when processed individually (Supplementary Fig. 5e) show that attempts at replication were successful.
	Cryo-EM: No replication studies were attempted, because cryo-EM and protein structure validation tools are sufficient to judge the quality of the structures.
	Growth assays (Figs. 5b-d; Extended Data Figs. 7 and 9; Supplementary Fig. 4b): Attempts on reproducing the growth assays in at least one additional biological replicate were successful.
	ATPase activity assays (Figs. 5h, 6h-j; Extended Data Fig. 1c): Attempts on reproducing the ATPase activity assays in a second biological replicate were successful.
	Fluorescence drug transport assays (Figs. 5e and f, 6e-g, Extended Data Fig. 10a): Attempts on reproducing the fluorescence transport assays in a second biological replicate were successful.
	EfrCD production measurements based on GFP (Figs. 5g, Extended Data Fig. 10b; Supplementary Fig. 4a): Attempts on reproducing the GFP fluorescence assays in a second biological replicate were successful.
Randomization	DMS: DMS libraries were grown as pools in culture flasks. No randomization was performed. The covariate of growth bias of variants was controlled by performing the DMS experiment in the absence of drugs (Extended Data Fig. 2), thereby removing 16 variants scores from our analysis.
	Cryo-EM: For model validation, the dataset was split into two independent randomized half sets.
	For growth assays, ATPase activity assays, fluorescence drug transport assays and GFP fluorescence measurements, no randomization of sample was performed, because we did not draw any conclusions from comparing groups of EfrCD variants. Instead individual variants of EfrCD were compared.

Blinding

DMS: The authors were not blinded. No blinding was required, because no subjective analyses were performed.

Cryo-EM: A cisTEM de-novo generated low-pass filtered initial model of EfrCD was used to avoid model bias.

For growth assays, ATPase activity assays, fluorescence drug transport assays and GFP fluorescence measurements, the authors were not blinded, because no subjective analyses were performed.

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

#### Materials & experimental systems Methods Involved in the study n/a Involved in the study n/a Antibodies $\mathbf{X}$ ChIP-seq $\boxtimes$ $\mathbf{X}$ Eukaryotic cell lines Flow cytometry $\mathbf{X}$ Palaeontology and archaeology MRI-based neuroimaging Animals and other organisms $\boxtimes$ Human research participants $\boxtimes$ Clinical data Dual use research of concern $\mathbf{X}$

# Antibodies

Antibodies used	Nanobody Nb_EfrCD#1 was generated as part of this study and is made available on request.	
Validation	The nanobody was validated by ELISA, its affinity was determined by GCI, its inhibition capacity was tested in an ATPase activity assay with purified EfrCD and it was used to determine the crvo-EM structure of EfrCD.	

# Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	A male 3 year old alpaca (Vicugna pacos) was used for immunization with EfrCD.	
Wild animals	No wild animals were used in this study.	
Field-collected samples	No field-collected samples were part of this study.	
Ethics oversight	Immunizations of alpacas were approved by the Cantonal Veterinary Office in Zurich, Switzerland (animal experiment licence nr. 188/2011)	

Note that full information on the approval of the study protocol must also be provided in the manuscript.